

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
25 October 2001 (25.10.2001)

PCT

(10) International Publication Number  
**WO 01/79473 A2**

(51) International Patent Classification<sup>2</sup>: C12N 9/00

(21) International Application Number: PCT/US01/40483

(22) International Filing Date: 11 April 2001 (11.04.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:  
60/197,508 18 April 2000 (18.04.2000) US

(63) Related by continuation (CON) or continuation-in-part (CIP) to earlier application:  
— US 60/197,508 (CIP)  
— Filed on 18 April 2000 (18.04.2000)

(71) Applicant (for all designated States except US): MILLENNIUM PHARMACEUTICALS, INC. [US/US]; 75 Sidney Street, Cambridge, MA 02139 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): MEYERS, Rachel, A. [US/US]; 115 Devonshire Road, Newton, MA 02468

(US). WILLIAMSON, Mark [US/US]; 15 Stonecrest Drive, Saugus, MA 01906 (US).

(74) Agents: MYERS, Louis et al.; Fish & Richardson P.C., 225 Franklin Street, Boston, MA 02110-2804 (US).

(81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:

— without international search report and to be republished upon receipt of that report

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.



A2  
WO 01/79473

(54) Title: 21953, A NOVEL HUMAN PROLYL OLIGOPEPTIDASE FAMILY MEMBER AND USES THEREOF

(57) Abstract: The invention provides isolated nucleic acids molecules, designated 21953 nucleic acid molecules, which encode novel prolyl oligopeptidase members. The invention also provides antisense nucleic acid molecules, recombinant expression vectors containing 21953 nucleic acid molecules, host cells into which the expression vectors have been introduced, and nonhuman transgenic animals in which a 21953 gene has been introduced or disrupted. The invention still further provides isolated 21953 proteins, fusion proteins, antigenic peptides and anti-21953 antibodies. Diagnostic methods utilizing compositions of the invention are also provided.

**21953, A NOVEL HUMAN PROLYL OLIGOPEPTIDASE FAMILY MEMBER  
AND USES THEREOF**

**Related Applications**

[01] This application claims priority to U.S. provisional application number 60/197,508 filed on April 18, 2000, the contents of which are incorporated herein by reference.

**Background of the Invention**

[02] Proline residues confer unique structural constraints on peptide chains and markedly influence the susceptibility of proximal peptide bonds to protease activity. For example, proline residues are sterically constrained by the imino group. Prolyl oligopeptidases are a distinct sub-group of endopeptidases that degrade a variety of proline-containing peptides by cleaving the peptide bond at the carboxyl side of proline residues. Some prolyl oligopeptidases prefer smaller polypeptides or oligopeptides as substrates.

[03] The natural substrates of prolyl oligopeptidases include many biologically active peptides such as peptide messenger molecules. For example, they are involved in the metabolism of peptide hormones and neuropeptides. Prolyl oligopeptidases have few naturally occurring inhibitors and their distinctive specificity prevents them from interacting with  $\beta$ -macroglobulin, unlike the great majority of endopeptidases. The specificity of an oligopeptidase depends on the three dimensional structure of its active site, which includes a putative catalytic triad, which contains aspartate, serine and histidine residues.

[04] Examples of known prolyl oligopeptidases include human prolyl oligopeptidase (Yoshimoto et al. Genebank AB020018), mouse prolyl oligopeptidase (Ishino et al., *J. Biochem.* 123 (3), 540-545 (1998)), pig prolyl oligopeptidase (Rennix et al., *Biochemistry*, 30:2195-2203, 1991), rat dipeptidyl-peptidase IV (Ognata et al., *J. Biol. Chem.*, 264:3596-3601, 1989), *F. meningosepticum* prolyl oligopeptidase (Yoshimoto et al., *J. Biochem.* 110:873-878, 1991), and *E. coli* protease II (Kanatani et al., *J. Biochemistry* (Tokyo), 110: 315-320, 1991).

[05] Prolyl oligopeptidases also control the activity of other peptides present in body fluids such as bradykinin and angiotensin. Bradykinin is a very potent vasodilator that increases the permeability of post capillary venules and acts on endothelial cells to activate phospholipase A2. Angiotensin causes contraction of vascular smooth muscle, raising blood

pressure and stimulating aldosterone release from the adrenal glands. Other members of the prolyl oligopeptidase family mediate the degradation of neuropeptides such as substance P, thyrotropin releasing hormone, hippocampal cholinergic neurostimulating peptide (HCNP), neuropeptide Y (NPY), and neuropeptides derived from pro-opiomelanocortin (POMC) and neurohypophyseal hormones.

#### Summary of the Invention

[06] The present invention is based, in part, on the discovery of a novel prolyl oligopeptidase family member, referred to herein as "21953". The nucleotide sequence of a cDNA encoding 21953 is shown in SEQ ID NO:1, and the amino acid sequence of a 21953 polypeptide is shown in SEQ ID NO:2. In addition, the nucleotide sequences of the coding region are depicted in SEQ ID NO:3.

[07] Accordingly, in one aspect, the invention features a nucleic acid molecule which encodes a 21953 protein or polypeptide, e.g., a biologically active portion of the 21953 protein. In a preferred embodiment the isolated nucleic acid molecule encodes a polypeptide having the amino acid sequence of SEQ ID NO:2. In other embodiments, the invention provides isolated 21953 nucleic acid molecules having the nucleotide sequence shown in SEQ ID NO:1, SEQ ID NO:3, or the sequence of the DNA insert of the plasmid deposited with ATCC Accession Number \_\_\_\_\_. In still other embodiments, the invention provides nucleic acid molecules that are substantially identical (e.g., naturally occurring allelic variants) to the nucleotide sequence shown in SEQ ID NO:1, SEQ ID NO:3, or the sequence of the DNA insert of the plasmid deposited with ATCC Accession Number \_\_\_\_\_. In other embodiments, the invention provides a nucleic acid molecule which hybridizes under a stringency condition described herein to a nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO:1, 3, or the sequence of the DNA insert of the plasmid deposited with ATCC Accession Number \_\_\_\_\_, wherein the nucleic acid encodes a full length 21953 protein or an active fragment thereof.

[08] In a related aspect, the invention further provides nucleic acid constructs which include a 21953 nucleic acid molecule described herein. In certain embodiments, the nucleic acid molecules of the invention are operatively linked to native or heterologous regulatory sequences. Also included, are vectors and host cells containing the 21953 nucleic acid molecules of the invention e.g., vectors and host cells suitable for producing 21953 nucleic acid molecules and polypeptides.

[09] In another related aspect, the invention provides nucleic acid fragments suitable as primers or hybridization probes for the detection of 21953-encoding nucleic acids.

[10] In still another related aspect, isolated nucleic acid molecules that are antisense to a 21953 encoding nucleic acid molecule are provided.

[11] In another aspect, the invention features, 21953 polypeptides, and biologically active or antigenic fragments thereof that are useful, e.g., as reagents or targets in assays applicable to treatment and diagnosis of 21953-mediated or -related disorders. In another embodiment, the invention provides 21953 polypeptides having a 21953 activity. Preferred polypeptides are 21953 proteins including at least one prolyl oligopeptidase domain, and, preferably, having a 21953 activity, e.g., a 21953 activity as described herein.

[12] In other embodiments, the invention provides 21953 polypeptides, e.g., a 21953 polypeptide having the amino acid sequence shown in SEQ ID NO:2 or the amino acid sequence encoded by the cDNA insert of the plasmid deposited with ATCC Accession Number \_\_\_\_; an amino acid sequence that is substantially identical to the amino acid sequence shown in SEQ ID NO:2 or the amino acid sequence encoded by the cDNA insert of the plasmid deposited with ATCC Accession Number \_\_\_\_; or an amino acid sequence encoded by a nucleic acid molecule having a nucleotide sequence which hybridizes under a stringency condition described herein to a nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO:1, SEQ ID NO:3, or the sequence of the DNA insert of the plasmid deposited with ATCC Accession Number \_\_\_\_ , wherein the nucleic acid encodes a full length 21953 protein or an active fragment thereof.

[13] In a related aspect, the invention further provides nucleic acid constructs which include a 21953 nucleic acid molecule described herein.

[14] In a related aspect, the invention provides 21953 polypeptides or fragments operatively linked to non-21953 polypeptides to form fusion proteins.

[15] In another aspect, the invention features antibodies and antigen-binding fragments thereof, that react with, or more preferably specifically bind 21953 polypeptides.

[16] In another aspect, the invention provides a method of evaluating a sample. The method includes: providing a sample; detecting a 21953 polypeptide or nucleic acid in the sample; and, optionally, comparing the level of expressed 21953 molecules to a reference sample. For example, an increased level of 21953 molecules can be an indication that the sample includes cells transiting from the G1 cell cycle phase to S phase. In other examples, the level of 21953 molecules can be an indication that a sample includes a proliferating cell,

e.g., a proliferating lung, breast, ovary, or colon cell; or a heart cell, a prostate cell, a vascular cell (e.g., a smooth muscle or an endothelial cell), or a brain cell.

[17] In another aspect, the invention provides methods of screening for compounds that modulate the expression or activity of the 21953 polypeptides or nucleic acids. The invention also provides assays for determining the activity of or the presence or absence of 21953 polypeptides or nucleic acid molecules in a biological sample, including for disease diagnosis.

[18] In still another aspect, the invention provides a process for modulating 21953 polypeptide or nucleic acid expression or activity, e.g. using the screened compounds. In certain embodiments, the methods involve treatment of conditions or disorders related to aberrant activity or expression of the 21953 polypeptides or nucleic acids, such as conditions or disorders involving aberrant or deficient cell proliferation or differentiation, e.g., a cancer (e.g. a cancer of the lung, breast, ovary, prostate, or colon), or conditions or disorders of the cardiovascular (including vascular, e.g., a smooth muscle or an endothelial cell), neuronal, or reproductive (e.g., prostatic) systems.

[19] Yet another aspect, the invention provides methods for modulating the activity of a 21953-expressing cell; e.g., a hyper-proliferative 21953-expressing cell. In one embodiment, the activity is modulated by one of more of: inhibiting the proliferation or migration or inducing the differentiation or killing of the 21953-expressing cell. The method includes contacting the cell with a compound (e.g., a compound identified using the methods described herein) that modulates the activity, or expression, of the 21953 polypeptide or nucleic acid, such that the activity of the 21953-expressing cell is modulated.

[20] In a preferred embodiment, the contacting step is effective *in vitro* or *ex vivo*. In other embodiments, the contacting step is effected *in vivo*, e.g., in a subject (e.g., a mammal, e.g., a human), as part of a therapeutic or prophylactic protocol.

[21] In a preferred embodiment, the cell is a hyperproliferative cell, e.g., a cell found in a solid tumor, a soft tissue tumor, or a metastatic lesion. For example, the cell is a lung, breast, ovary, prostate, or colon cell. In a preferred embodiment, the cell is lung cell.

[22] In other embodiments, the cell is a neural cell (e.g., a neuronal or a glial cell), a vascular cell (e.g., smooth muscle or an endothelial cell), a heart cell, a prostatic cell, or an immune cell.

[23] In a preferred embodiment, the compound is an inhibitor of a 21953 polypeptide. Preferably, the inhibitor is chosen from a peptide, a phosphopeptide, a small

organic molecule, a small inorganic molecule and an antibody (e.g., an antibody conjugated to a therapeutic moiety selected from a cytotoxin, a cytotoxic agent and a radioactive metal ion). The inhibitor can also be a protease inhibitor or a derivative thereof, or a peptidomimetic, e.g., a phosphonate analog of a peptide substrate such as a prolyl peptide substrate. In another preferred embodiment, the compound is an inhibitor of a 21953 nucleic acid, e.g., an antisense, a ribozyme, or a triple helix molecule.

[24] In a preferred embodiment, the compound is administered in combination with a cytotoxic agent. Examples of cytotoxic agents include anti-microtubule agent, a topoisomerase I inhibitor, a topoisomerase II inhibitor, an anti-metabolite, a mitotic inhibitor, an alkylating agent, an intercalating agent, an agent capable of interfering with a signal transduction pathway, an agent that promotes apoptosis or necrosis, and radiation.

[25] In another aspect, the invention features methods for treating or preventing a disorder characterized by aberrant activity, e.g., cellular proliferation or differentiation, of a 21953-expressing cell, in a subject. Preferably, the method includes comprising administering to the subject (e.g., a mammal, e.g., a human) an effective amount of a compound (e.g., a compound identified using the methods described herein) that modulates the activity, or expression, of the 21953 polypeptide or nucleic acid.

[26] In a preferred embodiment, the disorder is a cancerous or pre-cancerous condition, e.g., relating to proliferation of a lung, breast, ovary, prostate, or colon cell. In another preferred embodiment, the disorder is an immune, a neuronal, cardiovascular, reproductive disorder, e.g., a disorder relating to aberrant processing of a polypeptide hormone.

[27] In a further aspect, the invention provides methods for evaluating the efficacy of a treatment of a disorder, e.g., proliferative disorder (e.g., lung cancer), or a neuronal disorder. The method includes: treating a subject, e.g., a patient or an animal, with a protocol under evaluation (e.g., treating a subject with one or more of: chemotherapy, radiation, and/or a compound identified using the methods described herein); and evaluating the expression of a 21953 nucleic acid or polypeptide before and after treatment. A change, e.g., a decrease or increase, in the level of a 21953 nucleic acid (e.g., mRNA) or polypeptide after treatment, relative to the level of expression before treatment, is indicative of the efficacy of the treatment of the disorder. The level of 21953 nucleic acid or polypeptide expression can be detected, e.g., by a method described herein.

[28] In a preferred embodiment, the evaluating step includes obtaining a sample (e.g., a tissue sample, e.g., a biopsy, or a fluid sample) from the subject, before and after treatment

and comparing the level of expressing of a 21953 nucleic acid (e.g., mRNA) or polypeptide before and after treatment.

[29] In another aspect, the invention provides methods for evaluating the efficacy of a therapeutic or prophylactic agent (e.g., an anti-neoplastic agent). The method includes: contacting a sample with an agent (e.g., a compound identified using the methods described herein, a cytotoxic agent) and, evaluating the expression of 21953 nucleic acid or polypeptide in the sample before and after the contacting step. A change, e.g., a decrease or increase, in the level of 21953 nucleic acid (e.g., mRNA) or polypeptide in the sample obtained after the contacting step, relative to the level of expression in the sample before the contacting step, is indicative of the efficacy of the agent. The level of 21953 nucleic acid or polypeptide expression can be detected by any method described herein. In a preferred embodiment, the sample includes cells obtained from a cancerous, a neuronal, immune, a cardiovascular, or prostatic tissue. The cancerous tissue can include, for example, cells of lung, breast, ovary, prostate, or colon.

[30] In further aspect, the invention provides assays for determining the presence or absence of a genetic alteration in a 21953 polypeptide or nucleic acid molecule, including for disease diagnosis. In a still further aspect, the invention features a method of processing a polypeptide-hormone precursor, e.g., in vitro.

[31] In another aspect, the invention features a two dimensional array having a plurality of addresses, each address of the plurality being positionally distinguishable from each other address of the plurality, and each address of the plurality having a unique capture probe, e.g., a nucleic acid or peptide sequence. At least one address of the plurality has a capture probe that recognizes a 21953 molecule. In one embodiment, the capture probe is a nucleic acid, e.g., a probe complementary to a 21953 nucleic acid sequence. In another embodiment, the capture probe is a polypeptide, e.g., an antibody specific for 21953 polypeptides. Also featured is a method of analyzing a sample by contacting the sample to the aforementioned array and detecting binding of the sample to the array.

[32] Other features and advantages of the invention will be apparent from the following detailed description, and from the claims.

#### Brief Description of the Drawings

[33] *Figure 1* depicts a hydropathy plot of human 21953. Relative hydrophobic residues are shown above the dashed horizontal line, and relative hydrophilic residues are

below the dashed horizontal line. Numbers corresponding to positions in the amino acid sequence of human 21953 are indicated.

[34] *Figure 2* depicts an alignment of the prolyl oligopeptidase domain of human 21953 with a consensus amino acid sequence derived from a hidden Markov model for prolyl oligopeptidase domains. The upper sequence is the consensus amino acid sequence (SEQ ID NO:4), while the lower amino acid sequence corresponds to amino acids 672 to 744 of SEQ ID NO:2.

[35] *Figures 3A and 3B* depict an alignment of human dipeptidyl peptidase IV (Accession Number P48147) (upper line, SEQ ID NO:5), to the 21953 amino acid sequence. The \* symbol indicates identities, and the : or . symbols indicate similarities. The alignment was generated by ClustalW (Thompson et al. (1994) Nucleic Acids Res. 22:4673-4680).

#### Detailed Description

[36] The human 21953 sequence (see SEQ ID NO:1 as recited in Example 1), which is approximately 3143 nucleotides long, including untranslated regions, contains a predicted methionine-initiated coding sequence of about 2649 nucleotides, including the termination codon. The coding sequence encodes a 882 amino acid protein (see SEQ ID NO:2 as recited in Example 1).

[37] Human 21953 contains the following regions or other structural features: a predicted prolyl oligopeptidase domain (PFAM Accession PF00326) located at about amino acids 672-744 of SEQ ID NO:2; two predicted cAMP phosphorylation sites and cGMP-dependent protein kinase phosphorylation domains (Prosite Accession PS00004) located at about amino acid residues 231 to 234 of SEQ ID NO:2 and about amino acid residues 476-479 of SEQ ID NO:2; ten predicted Protein Kinase C sites (PS00005) at about amino acids 52 to 54, 80 to 82, 115 to 117, 307 to 309, 312 to 314, 326 to 328, 551 to 553, 594 to 596, 776 to 778, and 850 to 852 of SEQ ID NO:2; 11 predicted Casein Kinase II sites (PS00006) located at about amino 133 to 136, 227 to 230, 293 to 296, 412 to 415, 443 to 446, 499 to 502, 530 to 533, 587 to 590, 603 to 606, 615 to 618, and 723 to 726 of SEQ ID NO:2; five predicted tyrosine phosphorylation sites (PS00007) at about amino acids 29 to 36, 47 to 55, 308 to 315, 549 to 555, and 837 to 844 of SEQ ID NO:2; four predicted N-myristylation sites (PS00008) from about amino 176 to 181, 741 to 746, 762 to 767 and 873 to 878 of SEQ ID NO:2 and one predicted amidation site (PS00009) from about amino acid 642 to 645 of SEQ ID NO:2.

[38] For general information regarding PFAM identifiers, PS prefix and PF prefix domain identification numbers, refer to Sonnhammer et al. (1997) *Protein* 28:405-420 and <http://www.psc.edu/general/software/packages/pfam/pfam.html>.

[39] A plasmid containing the nucleotide sequence encoding human 21953 was deposited with American Type Culture Collection (ATCC), 10801 University Boulevard, Manassas, VA 20110-2209, on \_\_\_\_\_ and assigned Accession Number \_\_\_\_\_. This deposit will be maintained under the terms of the Budapest Treaty on the International Recognition of the Deposit of Microorganisms for the Purposes of Patent Procedure. This deposit was made merely as a convenience for those of skill in the art and is not an admission that a deposit is required under 35 U.S.C. §112.

[40] The 21953 polypeptide contains a significant number of structural characteristics in common with members of the human prolyl oligopeptidase family. The term "family" when referring to the protein and nucleic acid molecules of the invention means two or more proteins or nucleic acid molecules having a common structural domain or motif and having sufficient amino acid or nucleotide sequence homology as defined herein. Such family members can be naturally or non-naturally occurring and can be from either the same or different species. For example, a family can contain a first protein of human origin as well as other distinct proteins of human origin, or alternatively, can contain homologues of non-human origin, e.g., rat or mouse proteins. Members of a family can also have common functional characteristics.

[41] Polypeptide of the prolyl oligopeptidase family such as a 21953 polypeptide typically include an N-terminal seven-blade  $\beta$ -propeller domain and a C-terminal  $\alpha/\beta$  hydrolase domain. The N-terminal seven-blade  $\beta$ -propeller domain can include a "DPP IV N-terminal domain" or regions homologous with a "DPP IV N-terminal domain." The C-terminal  $\alpha/\beta$  hydrolase domain, e.g., the C-terminal region of a 21953 polypeptide, can include a "prolyl oligopeptidase domain" or regions homologous with a "prolyl oligopeptidase domain". The "prolyl oligopeptidase domain" can include a catalytic active site, which generally occurs at the C-terminal region of the polypeptide chain, which is involved in the hydrolysis of proline-containing peptide bonds. A prolyl oligopeptidase can be soluble. An alignment of human dipeptidyl peptidase IV (Accession Number P48147) to the 21953 amino acid sequence is depicted in Figures 3A-3B.

[42] As used herein, the term "prolyl oligopeptidase domain" includes an amino acid sequence of at least about 60 amino acid residues in length and having a bit score for the

alignment of the sequence to the Pfam Hidden Markov Model (HMM) PF00326 of at least 10. Preferably, a prolyl oligopeptidase domain includes at least about 30 to 180 amino acids, more preferably about 50 to 140 amino acid residues, or about 60 to 80 amino acids and has a bit score for the alignment of the sequence to the prolyl oligopeptidase domain (HMM) of at least 10, 20, 30 or greater. An alignment of the prolyl oligopeptidase domain (amino acids 672 to 744 of SEQ ID NO:2) of human 21953 with a consensus amino acid sequence derived from a hidden Markov model is depicted in Figure 3. In a preferred embodiment, a human 21953 polypeptide has a serine peptidase active site, e.g., an active site that is nearly identical to the Prosite signature PDOC00587. The active site can have a conserved catalytic triad with a conserved serine, e.g., a serine residue located at about amino acid 739 of SEQ ID NO:2, a conserved aspartic acid, e.g., an aspartic acid residue located at about amino acid 817 of SEQ ID NO:2, and a conserved histidine, e.g., a histidine residue located at about amino acid 849 of SEQ ID NO:2.

[43] In a preferred embodiment 21953 polypeptide or protein has a “prolyl oligopeptidase domain” or a region which includes at least about 30-300, more preferably about 50-150, or 60-80 amino acid residues and has at least about 50%, 60%, 70% 80% 90% 95%, 99%, or 100% homology with a “prolyl oligopeptidase domain,” e.g., the prolyl oligopeptidase domain of human 21953 (e.g., residues 672-744 of SEQ ID NO:2).

[44] To identify the presence of a “prolyl oligopeptidase” domain in a 21953 protein sequence, and make the determination that a polypeptide or protein of interest has a particular profile, the amino acid sequence of the protein can be searched against a database of HMMs (e.g., the Pfam database, release 2.1) using the default parameters ([http://www.sanger.ac.uk/Software/Pfam/HMM\\_search](http://www.sanger.ac.uk/Software/Pfam/HMM_search)). For example, the hmmsf program, which is available as part of the HMMER package of search programs, is a family specific default program for MILPAT0063 and a score of 15 is the default threshold score for determining a hit. Alternatively, the threshold score for determining a hit can be lowered (e.g., to 8 bits). A description of the Pfam database can be found in Sonhammer *et al.* (1997) *Proteins* 28(3):405-420 and a detailed description of HMMs can be found, for example, in Gribskov *et al.* (1990) *Meth. Enzymol.* 183:146-159; Gribskov *et al.* (1987) *Proc. Natl. Acad. Sci. USA* 84:4355-4358; Krogh *et al.* (1994) *J. Mol. Biol.* 235:1501-1531; and Stultz *et al.* (1993) *Protein Sci.* 2:305-314, the contents of which are incorporated herein by reference. A search was performed against the HMM database resulting in the identification

of a “prolyl oligopeptidase domain” domain in the amino acid sequence of human 21953 at about residues 672-744 of SEQ ID NO:2 (see Figure 2).

[45] In a preferred embodiment, a 21953 polypeptide includes an N-terminal seven-blade  $\beta$ -propeller domain, e.g., residues about 88 to 663 of SEQ ID NO:2. The amino acid sequence of this region can be aligned to the HMM profile for DPP IV N-terminal domain or the human DPP IV amino acid sequence (P27487). As used herein, the term “DPP IV N-terminal domain” refers to an amino acid sequence at least 60% identical to residues about 88 to 663 of SEQ ID NO:2.

[46] A 21953 family member can include a prolyl oligopeptidase domain and may also include a cAMP phosphorylation site and cGMP-dependent protein kinase phosphorylation domain, a predicted Protein Kinase C site, a predicted Casein Kinase II site, a predicted tyrosine phosphorylation site, a predicted N-myristylation site, and an amidation site.

[47] As the 21953 polypeptides of the invention may modulate 21953-mediated activities, e.g., a dipeptidyl peptidase activity such as a prolyl oligopeptidase activity, they may be useful for developing novel diagnostic and therapeutic agents for 21953-mediated or related disorders, as described below. The 21953 polypeptide of the invention are highly expressed in tumors, for example in breast and lung tumors. Further, 21953 polypeptide expression is increased at the G1/S phase transition of the mammalian cell cycle. Additional expression data for 21953 polypeptides are described below and in the Figures. Generally, increased prolyl oligopeptidase activity has been detected in human prostate, lung, and sigmoid tumors relative to healthy normal tissue. Such increased activity can result from 21953-increased expression.

[48] As used herein, a “21953 activity”, “biological activity of 21953” or “functional activity of 21953”, refers to an activity exerted by a 21953 protein, polypeptide or nucleic acid molecule on, e.g., a 21953-responsive cell or on a 21953 substrate, e.g., a oligopeptide substrate, as determined *in vivo* or *in vitro*. In one embodiment, a 21953 activity is a direct activity, such as an association with a 21953 target molecule. A “target molecule” or “binding partner” is a molecule with which a 21953 protein binds or interacts in nature. For example, the 21953 proteins of the present invention can have one or more of the following activities: (1) hydrolyzing peptide bonds at the carboxyl side of proline residues; (2) mediating degradation of proline-containing peptides, e.g., a prolyl endopeptidases activity; (3) processing of peptide factors (e.g., peptide hormones, chemokines, cytokines,

neuropeptides, and vasoactive peptides); (4) processing N-terminal dipeptides of unmodified N-termini wherein the penultimate residue is proline; (5) modulating cell proliferation and/or modulating cell differentiation (e.g., of a lung, breast, lymphoid, or colon cell); (6) modulating the regulation of transmission of intracellular signals, e.g., during immunological processes; (7) modulating metabolism of neurotransmitters or neuropeptides; (8) modulating neurodegeneration; or (9) modulating follicular development.

[49] As used herein, a “dipeptidyl peptidase activity” refers to a catalytic activity that accelerates the scission of a peptide bond between an amino acid sequence of less than four amino acids and the remainder of the polypeptide. Preferably, the cleaved peptide is a dipeptide having two amino acids. The catalytic activity can be mediated by the side chain of a serine amino acid and surrounding residues in the active site.

[50] As used herein, a “prolyl endopeptidases activity” refers to a catalytic activity that accelerates the scission of a peptide bond adjacent to a proline amino acid in a peptide or polypeptide chain. This catalytic activity has been detected, for example, in primary human lung tumors, squamous cell lung carcinomas, and lung adenocarcinomas. For example, squamous cell lung carcinomas and lung adenocarcinomas showed significantly higher levels of prolyl endopeptidases activity relative to normal lung parenchyma.

[51] In accordance with the above-described sequence similarities and observed polypeptide expression pattern, the 21953 molecules of the present invention can have similar biological activities as related prolyl oligopeptidase family members. Members of the prolyl oligopeptidase family can play an important role in the metabolism of a variety of proline containing peptides by cleaving prolyl bonds. These peptides can be less than about 200, 150, 100, or 50 residues in length. Prolyl oligopeptidases are involved, e.g., alone or together with other factors, in the regulation, e.g., processing, activation, or degradation of biological factors, e.g., peptide hormones (such as growth hormone, insulin, prolactin, adrenocorticotropic hormone, placental lactogen, calcitonin, parathyroid hormone, and thyroid stimulating hormone); chemokines; cytokines; neuropeptides; and vasoactive peptides.

[52] As the 21953 mRNA is highly expressed, for example, in cancerous tissues (e.g., lung and breast tumors), as well as normal cardiovascular, neural, and prostatic tissues, the molecules of the invention can be used to treat, prevent and/or diagnose disorders involving aberrant activity of 21953-expressing cells. Accordingly, the 21953 molecules can act as novel diagnostic targets and therapeutic agents for controlling disorders associated with the

aberrant activity or degradation of peptide hormones, e.g., disorders associated with cell differentiation and proliferation (e.g., a cancer of the lung, breast, ovary, and colon tissues), immune function (e.g., T cell activities, e.g., lymphomas, leukemias, and immune disorders), reproductive, neurological and cardiovascular function.

[53] As used herein, the terms "cancer", "hyperproliferative" and "neoplastic" refer to cells having the capacity for autonomous growth, i.e., an abnormal state or condition characterized by rapidly proliferating cell growth. Hyperproliferative and neoplastic disease states may be categorized as pathologic, i.e., characterizing or constituting a disease state, or may be categorized as non-pathologic, i.e., a deviation from normal but not associated with a disease state. The term is meant to include all types of cancerous growths or oncogenic processes, metastatic tissues or malignantly transformed cells, tissues, or organs, irrespective of histopathologic type or stage of invasiveness. "Pathologic hyperproliferative" cells occur in disease states characterized by malignant tumor growth. Examples of non-pathologic hyperproliferative cells include proliferation of cells associated with wound repair.

[54] Examples of cellular proliferative and/or differentiative disorders include cancer, e.g., carcinoma, sarcoma, or metastatic disorders. The 21953 molecules can act as novel diagnostic targets and therapeutic agents for controlling lung cancer, breast cancer, ovarian cancer, colon cancer, metastasis of such cancers and the like. A metastatic tumor can arise from a multitude of primary tumor types, including but not limited to those of lung, breast, liver, colon and ovarian origin.

[55] Examples of cellular proliferative and/or differentiative disorders of the lung include, but are not limited to, squamous cell lung carcinomas, small cell lung carcinoma, lung adenocarcinomas, bronchogenic carcinoma, including paraneoplastic syndromes, bronchioloalveolar carcinoma, neuroendocrine tumors, such as bronchial carcinoid, miscellaneous tumors, and metastatic tumors; pathologies of the pleura, including inflammatory pleural effusions, noninflammatory pleural effusions, pneumothorax, and pleural tumors, including solitary fibrous tumors (pleural fibroma) and malignant mesothelioma.

[56] Examples of cellular proliferative and/or differentiative disorders of the breast include, but are not limited to, proliferative breast disease including, e.g., epithelial hyperplasia, sclerosing adenosis, and small duct papillomas; tumors, e.g., stromal tumors such as fibroadenoma, phyllodes tumor, and sarcomas, and epithelial tumors such as large duct papilloma; carcinoma of the breast including in situ (noninvasive) carcinoma that

includes ductal carcinoma in situ (including Paget's disease) and lobular carcinoma in situ, and invasive (infiltrating) carcinoma including, but not limited to, invasive ductal carcinoma, invasive lobular carcinoma, medullary carcinoma, colloid (mucinous) carcinoma, tubular carcinoma, and invasive papillary carcinoma, and miscellaneous malignant neoplasms. Disorders in the male breast include, but are not limited to, gynecomastia and carcinoma.

[57] Examples of cellular proliferative and/or differentiative disorders of the colon include, but are not limited to, non-neoplastic polyps, adenomas, familial syndromes, colorectal carcinogenesis, colorectal carcinoma, and carcinoid tumors.

[58] Examples of cellular proliferative and/or differentiative disorders of the liver include, but are not limited to, nodular hyperplasias, adenomas, and malignant tumors, including primary carcinoma of the liver and metastatic tumors.

[59] Examples of cellular proliferative and/or differentiative disorders of the ovary include, but are not limited to, ovarian tumors such as, tumors of coelomic epithelium, serous tumors, mucinous tumors, endometrioid tumors, clear cell adenocarcinoma, cystadenofibroma, brenner tumor, surface epithelial tumors; germ cell tumors such as mature (benign) teratomas, monodermal teratomas, immature malignant teratomas, dysgerminoma, endodermal sinus tumor, choriocarcinoma; sex cord-stromal tumors such as, granulosa-theca cell tumors, thecoma-fibromas, androblastomas, hillock cell tumors, and gonadoblastoma; and metastatic tumors such as Krukenberg tumors.

[60] Additional examples of proliferative disorders include hematopoietic neoplastic disorders. As used herein, the term "hematopoietic neoplastic disorders" includes diseases involving hyperplastic/neoplastic cells of hematopoietic origin, e.g., arising from myeloid, lymphoid or erythroid lineages, or precursor cells thereof. Preferably, the diseases arise from poorly differentiated acute leukemias, e.g., erythroblastic leukemia and acute megakaryoblastic leukemia. Additional exemplary myeloid disorders include, but are not limited to, acute promyeloid leukemia (APML), acute myelogenous leukemia (AML) and chronic myelogenous leukemia (CML) (reviewed in Vaickus, L. (1991) *Crit Rev. in Oncol./Hematol.* 11:267-97); lymphoid malignancies include, but are not limited to acute lymphoblastic leukemia (ALL) which includes B-lineage ALL and T-lineage ALL, chronic lymphocytic leukemia (CLL), prolymphocytic leukemia (PLL), hairy cell leukemia (HLL) and Waldenstrom's macroglobulinemia (WM). Additional forms of malignant lymphomas include, but are not limited to non-Hodgkin lymphoma and variants thereof, peripheral T

cell lymphomas, adult T cell leukemia/lymphoma (ATL), cutaneous T-cell lymphoma (CTCL), large granular lymphocytic leukemia (LGF), Hodgkin's disease and Reed-Sternberg disease.

[61] The 21953 nucleic acid and protein of the invention can be used to treat and/or diagnose a variety of immune disorders, e.g., as a result of aberrant 21953 activity in T cells. Examples of immune disorders or diseases include, but are not limited to, autoimmune diseases (including, for example, diabetes mellitus, arthritis (including rheumatoid arthritis, juvenile rheumatoid arthritis, osteoarthritis, psoriatic arthritis), multiple sclerosis, encephalomyelitis, myasthenia gravis, systemic lupus erythematosus, autoimmune thyroiditis, dermatitis (including atopic dermatitis and eczematous dermatitis), psoriasis, Sjögren's Syndrome, Crohn's disease, aphthous ulcer, iritis, conjunctivitis, keratoconjunctivitis, ulcerative colitis, asthma, allergic asthma, cutaneous lupus erythematosus, scleroderma, vaginitis, proctitis, drug eruptions, leprosy reversal reactions, erythema nodosum leprosum, autoimmune uveitis, allergic encephalomyelitis, acute necrotizing hemorrhagic encephalopathy, idiopathic bilateral progressive sensorineural hearing loss, aplastic anemia, pure red cell anemia, idiopathic thrombocytopenia, polychondritis, Wegener's granulomatosis, chronic active hepatitis, Stevens-Johnson syndrome, idiopathic sprue, lichen planus, Graves' disease, sarcoidosis, primary biliary cirrhosis, uveitis posterior, and interstitial lung fibrosis), graft-versus-host disease, cases of transplantation, and allergy such as, atopic allergy.

[62] Examples of neuronal disorders include, but are not limited to disorders involving neurons, and disorders involving glia, such as astrocytes, oligodendrocytes, ependymal cells, and microglia; cerebral edema, raised intracranial pressure and herniation, and hydrocephalus; malformations and developmental diseases, such as neural tube defects, forebrain anomalies, posterior fossa anomalies, and syringomyelia and hydromyelia; perinatal brain injury; transmissible spongiform encephalopathies (prion diseases); demyelinating diseases, including multiple sclerosis, multiple sclerosis variants, acute disseminated encephalomyelitis and acute necrotizing hemorrhagic encephalomyelitis, and other diseases with demyelination; degenerative diseases, such as degenerative diseases affecting the cerebral cortex, including Alzheimer disease and Pick disease, degenerative diseases of basal ganglia and brain stem, including Parkinsonism, idiopathic Parkinson disease (paralysis agitans), progressive supranuclear palsy, corticobasal degeneration, multiple system atrophy, including striatonigral degeneration, Shy-Drager syndrome, and

olivopontocerebellar atrophy, and Huntington disease; spinocerebellar degenerations, including spinocerebellar ataxias, including Friedreich ataxia, and ataxia-telangiectasia, degenerative diseases affecting motor neurons, including amyotrophic lateral sclerosis (motor neuron disease), bulbospinal atrophy (Kennedy syndrome), and spinal muscular atrophy; tumors, such as gliomas, including astrocytoma, including fibrillary (diffuse) astrocytoma and glioblastoma multiforme, pilocytic astrocytoma, pleomorphic xanthoastrocytoma, and brain stem glioma, oligodendrogloma, and ependymoma and related paraventricular mass lesions, neuronal tumors, poorly differentiated neoplasms, including medulloblastoma, other parenchymal tumors, including primary brain lymphoma, germ cell tumors, and pineal parenchymal tumors, meningiomas, metastatic tumors, paraneoplastic syndromes, peripheral nerve sheath tumors, including schwannoma, neurofibroma, and malignant peripheral nerve sheath tumor (malignant schwannoma), and neurocutaneous syndromes (phakomatoses), including neurofibromatosis, including Type 1 neurofibromatosis (NF1) and TYPE 2 neurofibromatosis (NF2), tuberous sclerosis, and Von Hippel-Lindau disease.

[63] The term "vascular disorder" includes disorders involving aberrant activity (e.g., proliferation, metabolism, angiogenesis, vascularization) of blood vessel-associated cells, e.g., smooth muscle or endothelial cells. Examples of such disorders include but are not limited to hypertension (e.g., arterial hypertension), vascular restenosis, ischemic disease (e.g., atherosclerosis), tumorigenesis, tumor metastasis, diabetic retinopathy, endometriosis, Grave's disease. Aberrant vascular activity may also affect cardiovascular function, and thus the molecules of the invention can be used to treat, prevent and/or diagnose cardiovascular disorders. Examples of cardiovascular disorders, include but are not limited to, heart failure, cardiac hypertrophy, left-sided heart failure, and right-sided heart failure; ischemic heart disease, including but not limited to angina pectoris, myocardial infarction, chronic ischemic heart disease, and sudden cardiac death; hypertensive heart disease, including but not limited to, systemic (left-sided) hypertensive heart disease and pulmonary (right-sided) hypertensive heart disease; valvular heart disease, including but not limited to, valvular degeneration caused by calcification, such as calcific aortic stenosis, calcification of a congenitally bicuspid aortic valve, and mitral annular calcification, and myxomatous degeneration of the mitral valve (mitral valve prolapse), rheumatic fever and rheumatic heart disease, infective endocarditis, and noninfected vegetations, such as nonbacterial thrombotic endocarditis and endocarditis of systemic lupus erythematosus (Libman-Sacks disease),

carcinoid heart disease, and complications of artificial valves; myocardial disease, including but not limited to dilated cardiomyopathy, hypertrophic cardiomyopathy, restrictive cardiomyopathy, and myocarditis; pericardial disease, including but not limited to, pericardial effusion and hemopericardium and pericarditis, including acute pericarditis and healed pericarditis, and rheumatoid heart disease; neoplastic heart disease, including but not limited to, primary cardiac tumors, such as myxoma, lipoma, papillary fibroelastoma, rhabdomyoma, and sarcoma, and cardiac effects of noncardiac neoplasms; congenital heart disease, including but not limited to, left-to-right shunts--late cyanosis, such as atrial septal defect, ventricular septal defect, patent ductus arteriosus, and atrioventricular septal defect, right-to-left shunts--early cyanosis, such as tetralogy of fallot, transposition of great arteries, truncus arteriosus, tricuspid atresia, and total anomalous pulmonary venous connection, obstructive congenital anomalies, such as coarctation of aorta, pulmonary stenosis and atresia, and aortic stenosis and atresia, and disorders involving cardiac transplantation.

[64] As used herein, "a prostate disorder" refers to an abnormal condition occurring in the male pelvic region characterized by, e.g., male sexual dysfunction and/or urinary symptoms. This disorder may be manifested in the form of genitourinary inflammation (e.g., inflammation of smooth muscle cells) as in several common diseases of the prostate including prostatitis, benign prostatic hyperplasia and cancer, e.g., adenocarcinoma or carcinoma of the prostate.

[65] The 21953 nucleic acid and protein of the invention can be used to treat and/or diagnose a variety of conditions, in addition to the ones described above (see "Methods of Treatment" for additional examples).

[66] The presence of 21953 RNA or protein can also be used to identify a cell or tissue, or other biological sample, as being derived from breast, T-cell, kidney, liver, and aorta, or being of human origin. Expression can also be used to diagnose or stage a disorder, e.g., a cancer (e.g., a cancer of the lung or breast), or a breast, lymphoid, lung, ovarian, or liver disorder. Expression can be determined by evaluating RNA, e.g., by hybridization of a 21953 specific probe, or with a 21953 specific antibody.

[67] The 21953 protein, fragments thereof, and derivatives and other variants of the sequence in SEQ ID NO:2 thereof are collectively referred to as "polypeptides or proteins of the invention" or "21953 polypeptides or proteins". Nucleic acid molecules encoding such polypeptides or proteins are collectively referred to as "nucleic acids of the invention" or

“21953 nucleic acids.” 21953 molecules refer to 21953 nucleic acids, polypeptides, and antibodies.

[68] As used herein, the term “nucleic acid molecule” includes DNA molecules (e.g., a cDNA or genomic DNA), RNA molecules (e.g., an mRNA) and analogs of the DNA or RNA. A DNA or RNA analog can be synthesized from nucleotide analogs. The nucleic acid molecule can be single-stranded or double-stranded, but preferably is double-stranded DNA.

[69] The term “isolated nucleic acid molecule” or “purified nucleic acid molecule” includes nucleic acid molecules that are separated from other nucleic acid molecules present in the natural source of the nucleic acid. For example, with regards to genomic DNA, the term “isolated” includes nucleic acid molecules which are separated from the chromosome with which the genomic DNA is naturally associated. Preferably, an “isolated” nucleic acid is free of sequences which naturally flank the nucleic acid (i.e., sequences located at the 5' and/or 3' ends of the nucleic acid) in the genomic DNA of the organism from which the nucleic acid is derived. For example, in various embodiments, the isolated nucleic acid molecule can contain less than about 5 kb, 4 kb, 3 kb, 2 kb, 1 kb, 0.5 kb or 0.1 kb of 5' and/or 3' nucleotide sequences which naturally flank the nucleic acid molecule in genomic DNA of the cell from which the nucleic acid is derived. Moreover, an “isolated” nucleic acid molecule, such as a cDNA molecule, can be substantially free of other cellular material, or culture medium when produced by recombinant techniques, or substantially free of chemical precursors or other chemicals when chemically synthesized.

[70] As used herein, the term “hybridizes under low stringency, medium stringency, high stringency, or very high stringency conditions” describes conditions for hybridization and washing. Guidance for performing hybridization reactions can be found in *Current Protocols in Molecular Biology*, John Wiley & Sons, N.Y. (1989), 6.3.1-6.3.6, which is incorporated by reference. Aqueous and nonaqueous methods are described in that reference and either can be used. Specific hybridization conditions referred to herein are as follows: 1) low stringency hybridization conditions in 6X sodium chloride/sodium citrate (SSC) at about 45°C, followed by two washes in 0.2X SSC, 0.1% SDS at least at 50°C (the temperature of the washes can be increased to 55°C for low stringency conditions); 2) medium stringency hybridization conditions in 6X SSC at about 45°C, followed by one or more washes in 0.2X SSC, 0.1% SDS at 60°C; 3) high stringency hybridization conditions in 6X SSC at about 45°C, followed by one or more washes in 0.2X SSC, 0.1% SDS at 65°C;

and preferably 4) very high stringency hybridization conditions are 0.5M sodium phosphate, 7% SDS at 65°C, followed by one or more washes at 0.2X SSC, 1% SDS at 65°C. Very high stringency conditions (4) are the preferred conditions and the ones that should be used unless otherwise specified.

[71] Preferably, an isolated nucleic acid molecule of the invention that hybridizes under a stringency condition described herein to the sequence of SEQ ID NO:1 or SEQ ID NO:3, corresponds to a naturally-occurring nucleic acid molecule.

[72] As used herein, a "naturally-occurring" nucleic acid molecule refers to an RNA or DNA molecule having a nucleotide sequence that occurs in nature. For example a naturally occurring nucleic acid molecule can encode a natural protein.

[73] As used herein, the terms "gene" and "recombinant gene" refer to nucleic acid molecules which include at least an open reading frame encoding a 21953 protein. The gene can optionally further include non-coding sequences, e.g., regulatory sequences and introns. Preferably, a gene encodes a mammalian 21953 protein or derivative thereof.

[74] An "isolated" or "purified" polypeptide or protein is substantially free of cellular material or other contaminating proteins from the cell or tissue source from which the protein is derived, or substantially free from chemical precursors or other chemicals when chemically synthesized. "Substantially free" means that a preparation of 21953 protein is at least 10% pure. In a preferred embodiment, the preparation of 21953 protein has less than about 30%, 20%, 10% and more preferably 5% (by dry weight), of non-21953 protein (also referred to herein as a "contaminating protein"), or of chemical precursors or non-21953 chemicals. When the 21953 protein or biologically active portion thereof is recombinantly produced, it is also preferably substantially free of culture medium, i.e., culture medium represents less than about 20%, more preferably less than about 10%, and most preferably less than about 5% of the volume of the protein preparation. The invention includes isolated or purified preparations of at least 0.01, 0.1, 1.0, and 10 milligrams in dry weight.

[75] A "non-essential" amino acid residue is a residue that can be altered from the wild-type sequence of 21953 without abolishing or substantially altering a 21953 activity. Preferably the alteration does not substantially alter the 21953 activity, e.g., the activity is at least 20%, 40%, 60%, 70% or 80% of wild-type. An "essential" amino acid residue is a residue that, when altered from the wild-type sequence of 21953, results in abolishing a 21953 activity such that less than 20% of the wild-type activity is present. For example,

conserved amino acid residues in 21953 are predicted to be particularly unamenable to alteration.

[76] A “conservative amino acid substitution” is one in which the amino acid residue is replaced with an amino acid residue having a similar side chain. Families of amino acid residues having similar side chains have been defined in the art. These families include amino acids with basic side chains (e.g., lysine, arginine, histidine), acidic side chains (e.g., aspartic acid, glutamic acid), uncharged polar side chains (e.g., glycine, asparagine, glutamine, serine, threonine, tyrosine, cysteine), nonpolar side chains (e.g., alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan), beta-branched side chains (e.g., threonine, valine, isoleucine) and aromatic side chains (e.g., tyrosine, phenylalanine, tryptophan, histidine). Thus, a predicted nonessential amino acid residue in a 21953 protein is preferably replaced with another amino acid residue from the same side chain family. Alternatively, in another embodiment, mutations can be introduced randomly along all or part of a 21953 coding sequence, such as by saturation mutagenesis, and the resultant mutants can be screened for 21953 biological activity to identify mutants that retain activity. Following mutagenesis of SEQ ID NO:1 or SEQ ID NO:3, the encoded protein can be expressed recombinantly and the activity of the protein can be determined.

[77] As used herein, a “biologically active portion” of a 21953 protein includes a fragment of a 21953 protein which participates in an interaction, e.g., an intramolecular or an inter-molecular interaction. An inter-molecular interaction can be a specific binding interaction or an enzymatic interaction (e.g., the interaction can be transient and a covalent bond is formed or broken). An inter-molecular interaction can be between a 21953 molecule and a non-21953 molecule or between a first 21953 molecule and a second 21953 molecule (e.g., a dimerization interaction). Biologically active portions of a 21953 protein include peptides comprising amino acid sequences sufficiently homologous to or derived from the amino acid sequence of the 21953 protein, e.g., the amino acid sequence shown in SEQ ID NO:2, which include less amino acids than the full length 21953 proteins, and exhibit at least one activity of a 21953 protein. Typically, biologically active portions comprise a domain or motif with at least one activity of the 21953 protein, e.g., prolyl oligopeptidase activity. A biologically active portion of a 21953 protein can be a polypeptide which is, for example, 10, 25, 50, 100, 200 or more amino acids in length. Biologically active portions of a 21953 protein can be used as targets for developing agents which modulate a 21953 mediated activity, e.g., prolyl oligopeptidase activity.

[78] Calculations of homology or sequence identity between sequences (the terms are used interchangeably herein) are performed as follows.

[79] To determine the percent identity of two amino acid sequences, or of two nucleic acid sequences, the sequences are aligned for optimal comparison purposes (e.g., gaps can be introduced in one or both of a first and a second amino acid or nucleic acid sequence for optimal alignment and non-homologous sequences can be disregarded for comparison purposes). In a preferred embodiment, the length of a reference sequence aligned for comparison purposes is at least 30%, preferably at least 40%, more preferably at least 50%, 60%, and even more preferably at least 70%, 80%, 90%, 100% of the length of the reference sequence. The amino acid residues or nucleotides at corresponding amino acid positions or nucleotide positions are then compared. When a position in the first sequence is occupied by the same amino acid residue or nucleotide as the corresponding position in the second sequence, then the molecules are identical at that position (as used herein amino acid or nucleic acid "identity" is equivalent to amino acid or nucleic acid "homology").

[80] The percent identity between the two sequences is a function of the number of identical positions shared by the sequences, taking into account the number of gaps, and the length of each gap, which need to be introduced for optimal alignment of the two sequences.

[81] The comparison of sequences and determination of percent identity between two sequences can be accomplished using a mathematical algorithm. In a preferred embodiment, the percent identity between two amino acid sequences is determined using the Needleman and Wunsch ((1970) *J. Mol. Biol.* 48:444-453) algorithm which has been incorporated into the GAP program in the GCG software package (available at <http://www.gcg.com>), using either a Blossum 62 matrix or a PAM250 matrix, and a gap weight of 16, 14, 12, 10, 8, 6, or 4 and a length weight of 1, 2, 3, 4, 5, or 6. In yet another preferred embodiment, the percent identity between two nucleotide sequences is determined using the GAP program in the GCG software package (available at <http://www.gcg.com>), using a NWSgapdna.CMP matrix and a gap weight of 40, 50, 60, 70, or 80 and a length weight of 1, 2, 3, 4, 5, or 6. A particularly preferred set of parameters (and the one that should be used unless otherwise specified) are a Blossum 62 scoring matrix with a gap penalty of 12, a gap extend penalty of 4, and a frameshift gap penalty of 5.

[82] The percent identity between two amino acid or nucleotide sequences can be determined using the algorithm of E. Meyers and W. Miller ((1989) CABIOS, 4:11-17)

which has been incorporated into the ALIGN program (version 2.0), using a PAM120 weight residue table, a gap length penalty of 12 and a gap penalty of 4.

[83] The nucleic acid and protein sequences described herein can be used as a "query sequence" to perform a search against public databases to, for example, identify other family members or related sequences. Such searches can be performed using the NBLAST and XBLAST programs (version 2.0) of Altschul, *et al.* (1990) *J. Mol. Biol.* 215:403-10. BLAST nucleotide searches can be performed with the NBLAST program, score = 100, wordlength = 12 to obtain nucleotide sequences homologous to 21953 nucleic acid molecules of the invention. BLAST protein searches can be performed with the XBLAST program, score = 50, wordlength = 3 to obtain amino acid sequences homologous to 21953 protein molecules of the invention. To obtain gapped alignments for comparison purposes, Gapped BLAST can be utilized as described in Altschul *et al.*, (1997) *Nucleic Acids Res.* 25:3389-3402. When utilizing BLAST and Gapped BLAST programs, the default parameters of the respective programs (e.g., XBLAST and NBLAST) can be used. See <http://www.ncbi.nlm.nih.gov>.

[84] Particularly preferred 21953 polypeptides of the present invention have an amino acid sequence substantially identical to the amino acid sequence of SEQ ID NO:2. In the context of an amino acid sequence, the term "substantially identical" is used herein to refer to a first amino acid that contains a sufficient or minimum number of amino acid residues that are i) identical to, or ii) conservative substitutions of aligned amino acid residues in a second amino acid sequence such that the first and second amino acid sequences can have a common structural domain and/or common functional activity. For example, amino acid sequences that contain a common structural domain having at least about 60%, or 65% identity, likely 75% identity, more likely 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% identity to SEQ ID NO:2 are termed substantially identical.

[85] In the context of nucleotide sequence, the term "substantially identical" is used herein to refer to a first nucleic acid sequence that contains a sufficient or minimum number of nucleotides that are identical to aligned nucleotides in a second nucleic acid sequence such that the first and second nucleotide sequences encode a polypeptide having common functional activity, or encode a common structural polypeptide domain or a common functional polypeptide activity. For example, nucleotide sequences having at least about 60%, or 65% identity, likely 75% identity, more likely 85%, 90%, 91%, 92%, 93%, 94%,

95%, 96%, 97%, 98% or 99% identity to SEQ ID NO:1 or 3 are termed substantially identical.

[86] “Misexpression or aberrant expression”, as used herein, refers to a non-wildtype pattern of gene expression at the RNA or protein level. It includes: expression at non-wild type levels, i.e., over- or under-expression; a pattern of expression that differs from wild type in terms of the time or stage at which the gene is expressed, e.g., increased or decreased expression (as compared with wild type) at a predetermined developmental period or stage; a pattern of expression that differs from wild type in terms of altered, e.g., increased or decreased, expression (as compared with wild type) in a predetermined cell type or tissue type; a pattern of expression that differs from wild type in terms of the splicing size, translated amino acid sequence, post-transitional modification, or biological activity of the expressed polypeptide; a pattern of expression that differs from wild type in terms of the effect of an environmental stimulus or extracellular stimulus on expression of the gene, e.g., a pattern of increased or decreased expression (as compared with wild type) in the presence of an increase or decrease in the strength of the stimulus.

[87] “Subject,” as used herein, refers to human and non-human animals. The term “non-human animals” of the invention includes all vertebrates, e.g., mammals, such as non-human primates (particularly higher primates), sheep, dog, rodent (e.g., mouse or rat), guinea pig, goat, pig, cat, rabbits, cow, and non-mammals, such as chickens, amphibians, reptiles, etc. In a preferred embodiment, the subject is a human. In another embodiment, the subject is an experimental animal or animal suitable as a disease model.

[88] A “purified preparation of cells”, as used herein, refers to an in vitro preparation of cells. In the case of cells from multicellular organisms (e.g., plants and animals), a purified preparation of cells is a subset of cells obtained from the organism, not the entire intact organism. In the case of unicellular microorganisms (e.g., cultured cells and microbial cells), it consists of a preparation of at least 10% and more preferably 50% of the subject cells.

[89] Various aspects of the invention are described in further detail below.

#### Isolated Nucleic Acid Molecules

[90] In one aspect, the invention provides, an isolated or purified, nucleic acid molecule that encodes a 21953 polypeptide described herein, e.g., a full-length 21953 protein or a fragment thereof, e.g., a biologically active portion of 21953 protein. Also

included is a nucleic acid fragment suitable for use as a hybridization probe, which can be used, e.g., to identify a nucleic acid molecule encoding a polypeptide of the invention, 21953 mRNA, and fragments suitable for use as primers, e.g., PCR primers for the amplification or mutation of nucleic acid molecules.

[91] In one embodiment, an isolated nucleic acid molecule of the invention includes the nucleotide sequence shown in SEQ ID NO:1, or a portion of any of these nucleotide sequences. In one embodiment, the nucleic acid molecule includes sequences encoding the human 21953 protein (i.e., "the coding region" of SEQ ID NO:1, as shown in SEQ ID NO:3), as well as 5' untranslated sequences. Alternatively, the nucleic acid molecule can include only the coding region of SEQ ID NO:1 (e.g., SEQ ID NO:3) and, e.g., no flanking sequences which normally accompany the subject sequence. In another embodiment, the nucleic acid molecule encodes a sequence corresponding to a fragment of the protein that includes amino acid 672 to 744, 88 to 663, or 88 to 744 of SEQ ID NO:2.

[92] In another embodiment, an isolated nucleic acid molecule of the invention includes a nucleic acid molecule which is a complement of the nucleotide sequence shown in SEQ ID NO:1 or SEQ ID NO:3, or a portion of any of these nucleotide sequences. In other embodiments, the nucleic acid molecule of the invention is sufficiently complementary to the nucleotide sequence shown in SEQ ID NO:1 or SEQ ID NO:3, such that it can hybridize (e.g., under a stringency condition described herein) to the nucleotide sequence shown in SEQ ID NO:1 or 3, thereby forming a stable duplex.

[93] In one embodiment, an isolated nucleic acid molecule of the present invention includes a nucleotide sequence which is at least about 60%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more homologous to the entire length of the nucleotide sequence shown in SEQ ID NO:1 or SEQ ID NO:3, or a portion, preferably of the same length, of any of these nucleotide sequences.

### **21953 Nucleic Acid Fragments**

[94] A nucleic acid molecule of the invention can include only a portion of the nucleic acid sequence of SEQ ID NO:1 or 3. For example, such a nucleic acid molecule can include a fragment which can be used as a probe or primer or a fragment encoding a portion of a 21953 protein, e.g., an immunogenic or biologically active portion of a 21953 protein. A fragment can comprise those nucleotides of SEQ ID NO:1, which encode a prolyl oligopeptidase domain of human 21953. The nucleotide sequence determined from the

cloning of the 21953 gene allows for the generation of probes and primers designed for use in identifying and/or cloning other 21953 family members, or fragments thereof, as well as 21953 homologues, or fragments thereof, from other species.

[95] In another embodiment, a nucleic acid includes a nucleotide sequence that includes part, or all, of the coding region and extends into either (or both) the 5' or 3' noncoding region. Other embodiments include a fragment which includes a nucleotide sequence encoding an amino acid fragment described herein. Nucleic acid fragments can encode a specific domain or site described herein or fragments thereof, particularly fragments thereof which are at least 100, 150, 200, 300, 360, 400, 600, 650, or 700 amino acids in length. Fragments also include nucleic acid sequences corresponding to specific amino acid sequences described above or fragments thereof. Nucleic acid fragments should not to be construed as encompassing those fragments that may have been disclosed prior to the invention.

[96] A nucleic acid fragment can include a sequence corresponding to a domain, region, or functional site described herein. A nucleic acid fragment can also include one or more domain, region, or functional site described herein. Thus, for example, a 21953 nucleic acid fragment can include a sequence corresponding to a prolyl oligopeptidase domain.

[97] 21953 probes and primers are provided. Typically a probe/primer is an isolated or purified oligonucleotide. The oligonucleotide typically includes a region of nucleotide sequence that hybridizes under a stringency condition described herein to at least about 7, 12 or 15, preferably about 20 or 25, more preferably about 30, 35, 40, 45, 50, 55, 60, 65, or 75 consecutive nucleotides of a sense or antisense sequence of SEQ ID NO:1 or SEQ ID NO:3, or of a naturally occurring allelic variant or mutant of SEQ ID NO:1 or SEQ ID NO:3.

[98] In a preferred embodiment the nucleic acid is a probe which is at least 5 or 10, and less than 200, more preferably less than 100, or less than 50, base pairs in length. It should be identical, or differ by 1, or less than in 5 or 10 bases, from a sequence disclosed herein. If alignment is needed for this comparison the sequences should be aligned for maximum homology. "Looped" out sequences from deletions or insertions, or mismatches, are considered differences.

[99] A probe or primer can be derived from the sense or anti-sense strand of a nucleic acid which encodes: a fragment of the protein that includes amino acid 672 to 744, 88 to 663, or 88 to 744 of SEQ ID NO:2.

[100] In another embodiment a set of primers is provided, e.g., primers suitable for use in a PCR, which can be used to amplify a selected region of a 21953 sequence, e.g., a

domain, region, site or other sequence described herein. The primers should be at least 5, 10, or 50 base pairs in length and less than 100, or less than 200, base pairs in length. The primers should be identical, or differs by one base from a sequence disclosed herein or from a naturally occurring variant. For example, primers suitable for amplifying all or a portion of any of the following regions are provided: a prolyl oligopeptidase domain from about amino acid 672 to 744, 88 to 663, or 88 to 744 of SEQ ID NO:2.

[101] A nucleic acid fragment can encode an epitope bearing region of a polypeptide described herein.

[102] A nucleic acid fragment encoding a "biologically active portion of a 21953 polypeptide" can be prepared by isolating a portion of the nucleotide sequence of SEQ ID NO:1 or 3, which encodes a polypeptide having a 21953 biological activity (e.g., the biological activities of the 21953 proteins are described herein), expressing the encoded portion of the 21953 protein (e.g., by recombinant expression *in vitro*) and assessing the activity of the encoded portion of the 21953 protein. For example, a nucleic acid fragment encoding a biologically active portion of 21953 includes a prolyl oligopeptidase domain, e.g., amino acid residues about 672 to 744, 88 to 663, or 88 to 744 of SEQ ID NO:2. A nucleic acid fragment encoding a biologically active portion of a 21953 polypeptide, may comprise a nucleotide sequence which is greater than 361, 470, 800, 1000, 1600, or more nucleotides in length.

[103] In preferred embodiments, a nucleic acid includes a nucleotide sequence which is about 300, 400, 500, 700, 800, 1000, 1100, 1200, 1500, 1600, 2000, 2400 or more nucleotides in length and hybridizes under a stringency condition described herein to a nucleic acid molecule of SEQ ID NO:1, or SEQ ID NO:3. In a preferred embodiment, the nucleic acid includes a contiguous sequence that includes approximately nucleotide 1640, or 1642 of SEQ ID NO:1, e.g., the region from nucleotide 1635 to 1645 of SEQ ID NO:1. In other embodiment the nucleic acid includes a contiguous sequence that includes about nucleotides 1 to 25, 1 to 66, 100 to 300, 300 to 700, 500 to 800, 800 to 1200, 1000 to 1400, or 1200 to 1600 of SEQ ID NO:1.

#### 21953 Nucleic Acid Variants

[104] The invention further encompasses nucleic acid molecules that differ from the nucleotide sequence shown in SEQ ID NO:1 or SEQ ID NO:3. Such differences can be due to degeneracy of the genetic code (and result in a nucleic acid which encodes the same

21953 proteins as those encoded by the nucleotide sequence disclosed herein. In another embodiment, an isolated nucleic acid molecule of the invention has a nucleotide sequence encoding a protein having an amino acid sequence which differs, by at least 1, but less than 5, 10, 20, 50, or 100 amino acid residues that shown in SEQ ID NO:2. If alignment is needed for this comparison the sequences should be aligned for maximum homology. "Looped" out sequences from deletions or insertions, or mismatches, are considered differences.

[105] Nucleic acids of the inventor can be chosen for having codons, which are preferred, or non-preferred, for a particular expression system. E.g., the nucleic acid can be one in which at least one codon, at preferably at least 10%, or 20% of the codons has been altered such that the sequence is optimized for expression in *E. coli*, yeast, human, insect, or CHO cells.

[106] Nucleic acid variants can be naturally occurring, such as allelic variants (same locus), homologs (different locus), and orthologs (different organism) or can be non naturally occurring. Non-naturally occurring variants can be made by mutagenesis techniques, including those applied to polynucleotides, cells, or organisms. The variants can contain nucleotide substitutions, deletions, inversions and insertions. Variation can occur in either or both the coding and non-coding regions. The variations can produce both conservative and non-conservative amino acid substitutions (as compared in the encoded product).

[107] In a preferred embodiment, the nucleic acid differs from that of SEQ ID NO:1 or 3, e.g., as follows: by at least one but less than 10, 20, 30, or 40 nucleotides; at least one but less than 1%, 5%, 10% or 20% of the nucleotides in the subject nucleic acid. If necessary for this analysis the sequences should be aligned for maximum homology. "Looped" out sequences from deletions or insertions, or mismatches, are considered differences.

[108] Orthologs, homologs, and allelic variants can be identified using methods known in the art. These variants comprise a nucleotide sequence encoding a polypeptide that is 50%, at least about 55%, typically at least about 70-75%, more typically at least about 80-85%, and most typically at least about 90-95% or more identical to the nucleotide sequence shown in SEQ ID NO:2 or a fragment of this sequence. Such nucleic acid molecules can readily be identified as being able to hybridize under a stringency condition described herein, to the nucleotide sequence shown in SEQ ID NO 2 or a fragment of the sequence. Nucleic acid molecules corresponding to orthologs, homologs, and allelic variants of the 21953 cDNAs of

the invention can further be isolated by mapping to the same chromosome or locus as the 21953 gene.

[109] Preferred variants include those that are correlated with dipeptidyl peptidase or prolyl endopeptidases activity.

[110] Allelic variants of 21953, e.g., human 21953, include both functional and non-functional proteins. Functional allelic variants are naturally occurring amino acid sequence variants of the 21953 protein within a population that maintain the ability to bind and/or cleave polypeptide substrates, e.g., a polypeptide having a proline residue. Functional allelic variants will typically contain only conservative substitution of one or more amino acids of SEQ ID NO:2, or substitution, deletion or insertion of non-critical residues in non-critical regions of the protein. Non-functional allelic variants are naturally-occurring amino acid sequence variants of the 21953, e.g., human 21953, protein within a population that do not have the ability to bind and/or cleave polypeptide substrates, e.g., a polypeptide having a proline residue. Non-functional allelic variants will typically contain a non-conservative substitution, a deletion, or insertion, or premature truncation of the amino acid sequence of SEQ ID NO:2, or a substitution, insertion, or deletion in critical residues or critical regions of the protein.

[111] Moreover, nucleic acid molecules encoding other 21953 family members and, thus, which have a nucleotide sequence which differs from the 21953 sequences of SEQ ID NO:1 or SEQ ID NO:3 are intended to be within the scope of the invention.

**Antisense Nucleic Acid Molecules, Ribozymes and Modified 21953 Nucleic Acid Molecules**

[112] In another aspect, the invention features, an isolated nucleic acid molecule which is antisense to 21953. An "antisense" nucleic acid can include a nucleotide sequence which is complementary to a "sense" nucleic acid encoding a protein, e.g., complementary to the coding strand of a double-stranded cDNA molecule or complementary to an mRNA sequence. The antisense nucleic acid can be complementary to an entire 21953 coding strand, or to only a portion thereof (e.g., the coding region of human 21953 corresponding to SEQ ID NO:3). In another embodiment, the antisense nucleic acid molecule is antisense to a "noncoding region" of the coding strand of a nucleotide sequence encoding 21953 (e.g., the 5' and 3' untranslated regions).

[113] An antisense nucleic acid can be designed such that it is complementary to the entire coding region of 21953 mRNA, but more preferably is an oligonucleotide which is antisense to only a portion of the coding or noncoding region of 21953 mRNA. For example, the antisense oligonucleotide can be complementary to the region surrounding the translation start site of 21953 mRNA, e.g., between the -10 and +10 regions of the target gene nucleotide sequence of interest. An antisense oligonucleotide can be, for example, about 7, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, or more nucleotides in length.

[114] An antisense nucleic acid of the invention can be constructed using chemical synthesis and enzymatic ligation reactions using procedures known in the art. For example, an antisense nucleic acid (e.g., an antisense oligonucleotide) can be chemically synthesized using naturally occurring nucleotides or variously modified nucleotides designed to increase the biological stability of the molecules or to increase the physical stability of the duplex formed between the antisense and sense nucleic acids, e.g., phosphorothioate derivatives and acridine substituted nucleotides can be used. The antisense nucleic acid also can be produced biologically using an expression vector into which a nucleic acid has been subcloned in an antisense orientation (i.e., RNA transcribed from the inserted nucleic acid will be of an antisense orientation to a target nucleic acid of interest, described further in the following subsection).

[115] The antisense nucleic acid molecules of the invention are typically administered to a subject (e.g., by direct injection at a tissue site), or generated *in situ* such that they hybridize with or bind to cellular mRNA and/or genomic DNA encoding a 21953 protein to thereby inhibit expression of the protein, e.g., by inhibiting transcription and/or translation. Alternatively, antisense nucleic acid molecules can be modified to target selected cells and then administered systemically. For systemic administration, antisense molecules can be modified such that they specifically bind to receptors or antigens expressed on a selected cell surface, e.g., by linking the antisense nucleic acid molecules to peptides or antibodies which bind to cell surface receptors or antigens. The antisense nucleic acid molecules can also be delivered to cells using the vectors described herein. To achieve sufficient intracellular concentrations of the antisense molecules, vector constructs in which the antisense nucleic acid molecule is placed under the control of a strong pol II or pol III promoter are preferred.

[116] In yet another embodiment, the antisense nucleic acid molecule of the invention is an  $\alpha$ -anomeric nucleic acid molecule. An  $\alpha$ -anomeric nucleic acid molecule forms specific double-stranded hybrids with complementary RNA in which, contrary to the usual  $\beta$ -units, the strands run parallel to each other (Gaultier *et al.* (1987) *Nucleic Acids Res.* 15:6625-6641). The antisense nucleic acid molecule can also comprise a 2'-o-methylribonucleotide (Inoue *et al.* (1987) *Nucleic Acids Res.* 15:6131-6148) or a chimeric RNA-DNA analogue (Inoue *et al.* (1987) *FEBS Lett.* 215:327-330).

[117] In still another embodiment, an antisense nucleic acid of the invention is a ribozyme. A ribozyme having specificity for a 21953-encoding nucleic acid can include one or more sequences complementary to the nucleotide sequence of a 21953 cDNA disclosed herein (i.e., SEQ ID NO:1 or SEQ ID NO:3), and a sequence having known catalytic sequence responsible for mRNA cleavage (see U.S. Pat. No. 5,093,246 or Haselhoff and Gerlach (1988) *Nature* 334:585-591). For example, a derivative of a *Tetrahymena* L-19 IVS RNA can be constructed in which the nucleotide sequence of the active site is complementary to the nucleotide sequence to be cleaved in a 21953-encoding mRNA. See, e.g., Cech *et al.* U.S. Patent No. 4,987,071; and Cech *et al.* U.S. Patent No. 5,116,742. Alternatively, 21953 mRNA can be used to select a catalytic RNA having a specific ribonuclease activity from a pool of RNA molecules. See, e.g., Bartel, D. and Szostak, J.W. (1993) *Science* 261:1411-1418.

[118] 21953 gene expression can be inhibited by targeting nucleotide sequences complementary to the regulatory region of the 21953 (e.g., the 21953 promoter and/or enhancers) to form triple helical structures that prevent transcription of the 21953 gene in target cells. See generally, Helene, C. (1991) *Anticancer Drug Des.* 6:569-84; Helene, C. i (1992) *Ann. N.Y. Acad. Sci.* 660:27-36; and Maher, L.J. (1992) *Bioassays* 14:807-15. The potential sequences that can be targeted for triple helix formation can be increased by creating a so-called "switchback" nucleic acid molecule. Switchback molecules are synthesized in an alternating 5'-3', 3'-5' manner, such that they base pair with first one strand of a duplex and then the other, eliminating the necessity for a sizeable stretch of either purines or pyrimidines to be present on one strand of a duplex.

[119] The invention also provides detectably labeled oligonucleotide primer and probe molecules. Typically, such labels are chemiluminescent, fluorescent, radioactive, or colorimetric.

[120] A 21953 nucleic acid molecule can be modified at the base moiety, sugar moiety or phosphate backbone to improve, e.g., the stability, hybridization, or solubility of the molecule. For non-limiting examples of synthetic oligonucleotides with modifications see Toulmé (2001) *Nature Biotech.* 19:17 and Faria *et al.* (2001) *Nature Biotech.* 19:40-44. Such phosphoramidite oligonucleotides can be effective antisense agents.

[121] For example, the deoxyribose phosphate backbone of the nucleic acid molecules can be modified to generate peptide nucleic acids (see Hyrup B. *et al.* (1996) *Bioorganic & Medicinal Chemistry* 4: 5-23). As used herein, the terms "peptide nucleic acid" or "PNA" refers to a nucleic acid mimic, e.g., a DNA mimic, in which the deoxyribose phosphate backbone is replaced by a pseudopeptide backbone and only the four natural nucleobases are retained. The neutral backbone of a PNA can allow for specific hybridization to DNA and RNA under conditions of low ionic strength. The synthesis of PNA oligomers can be performed using standard solid phase peptide synthesis protocols as described in Hyrup B. *et al.* (1996) *supra* and Perry-O'Keefe *et al.* *Proc. Natl. Acad. Sci.* 93: 14670-675.

[122] PNAs of 21953 nucleic acid molecules can be used in therapeutic and diagnostic applications. For example, PNAs can be used as antisense or antigene agents for sequence-specific modulation of gene expression by, for example, inducing transcription or translation arrest or inhibiting replication. PNAs of 21953 nucleic acid molecules can also be used in the analysis of single base-pair mutations in a gene, (e.g., by PNA-directed PCR clamping); as 'artificial restriction enzymes' when used in combination with other enzymes, (e.g., S1 nucleases (Hyrup B. *et al.* (1996) *supra*)); or as probes or primers for DNA sequencing or hybridization (Hyrup B. *et al.* (1996) *supra*; Perry-O'Keefe *supra*).

[123] In other embodiments, the oligonucleotide may include other appended groups such as peptides (e.g., for targeting host cell receptors *in vivo*), or agents facilitating transport across the cell membrane (see, e.g., Letsinger *et al.* (1989) *Proc. Natl. Acad. Sci. USA* 86:6553-6556; Lemaitre *et al.* (1987) *Proc. Natl. Acad. Sci. USA* 84:648-652; PCT Publication No. W088/09810) or the blood-brain barrier (see, e.g., PCT Publication No. W089/10134). In addition, oligonucleotides can be modified with hybridization-triggered cleavage agents (see, e.g., Krol *et al.* (1988) *Bio-Techniques* 6:958-976) or intercalating agents (see, e.g., Zon (1988) *Pharm. Res.* 5:539-549). To this end, the oligonucleotide may be conjugated to another molecule, (e.g., a peptide, hybridization triggered cross-linking agent, transport agent, or hybridization-triggered cleavage agent).

[124] The invention also includes molecular beacon oligonucleotide primer and probe molecules having at least one region which is complementary to a 21953 nucleic acid of the invention, two complementary regions one having a fluorophore and one a quencher such that the molecular beacon is useful for quantitating the presence of the 21953 nucleic acid of the invention in a sample. Molecular beacon nucleic acids are described, for example, in Lizardi *et al.*, U.S. Patent No. 5,854,033; Nazarenko *et al.*, U.S. Patent No. 5,866,336, and Livak *et al.*, U.S. Patent 5,876,930.

#### Isolated 21953 Polypeptides

[125] In another aspect, the invention features, an isolated 21953 protein, or fragment, e.g., a biologically active portion, for use as immunogens or antigens to raise or test (or more generally to bind) anti-21953 antibodies. 21953 protein can be isolated from cells or tissue sources using standard protein purification techniques. 21953 protein or fragments thereof can be produced by recombinant DNA techniques or synthesized chemically.

[126] Polypeptides of the invention include those which arise as a result of the existence of multiple genes, alternative transcription events, alternative RNA splicing events, and alternative translational and post-translational events. The polypeptide can be expressed in systems, e.g., cultured cells, which result in substantially the same post-translational modifications present when expressed the polypeptide is expressed in a native cell, or in systems which result in the alteration or omission of post-translational modifications, e.g., glycosylation or cleavage, present when expressed in a native cell.

[127] In a preferred embodiment, a 21953 polypeptide has one or more of the following characteristics:

[128] (i) it has the ability to promote the degradation of proline-containing peptides by cleaving the peptide bond at the carboxyl side of proline residues;

[129] (ii) it has a molecular weight, (e.g., about 97 KDa), amino acid composition, or other physical characteristic, of a 21953 polypeptide, e.g., a polypeptide of SEQ ID NO:2;

[130] (iii) it has an overall sequence similarity of at least 60%, more preferably at least 70, 80, 90, or 95%, with a polypeptide of SEQ ID NO:2;

[131] (iv) it has a prolyl oligopeptidase domain which has preferably about 70%, 80%, 90% or 95% sequence similarity with amino acid residues 672-744 of SEQ ID NO:2; or

[132] (v) it has at least 70%, preferably 80%, and most preferably 90% of the cysteines found in the amino acid sequence of the native protein (SEQ ID NO:2).

[133] In a preferred embodiment the 21953 protein, or fragment thereof, differs from the corresponding sequence in SEQ ID:2. In one embodiment it differs by at least one but by less than 15, 10 or 5 amino acid residues. In another it differs from the corresponding sequence in SEQ ID NO:2 by at least one residue but less than 20%, 15%, 10% or 5% of the residues in it differ from the corresponding sequence in SEQ ID NO:2. (If this comparison requires alignment the sequences should be aligned for maximum homology. "Looped" out sequences from deletions or insertions, or mismatches, are considered differences.) The differences are, preferably, differences or changes at a non-essential residue or a conservative substitution. In a preferred embodiment the differences are not in the prolyl oligopeptidase domain and/or the DPP IV N-terminal domain. In another preferred embodiment one or more differences are in the prolyl oligopeptidase domain and/or the DPP IV N-terminal domain.

[134] Other embodiments include a protein that contain one or more changes in amino acid sequence, e.g., a change in an amino acid residue which is not essential for activity. Such 21953 proteins differ in amino acid sequence from SEQ ID NO:2, yet retain biological activity.

[135] In some embodiments, the protein includes an amino acid sequence at least about 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% or more homologous to SEQ ID NO:2. In some embodiments, the protein includes at least one contiguous amino acid from the region of about amino acid 1 to 200, 100 to 300, 200 to 400, 300 to 500, 400 to 600, 500 to 700, or 600 to 800 of SEQ ID NO:2.

[136] A 21953 protein or fragment is provided which varies from the sequence of SEQ ID NO:2 in regions defined by amino acids about 672 to 744 by at least one but by less than 15, 10 or 5 amino acid residues in the protein or fragment but which does not differ from SEQ ID NO:2 in regions defined by amino acids about 672 to 744. (If this comparison requires alignment the sequences should be aligned for maximum homology. "Looped" out sequences from deletions or insertions, or mismatches, are considered differences.) In some embodiments the difference is at a non-essential residue or is a conservative substitution, while in others the difference is at an essential residue or is a non-conservative substitution.

[137] In one embodiment, a biologically active portion of a 21953 protein includes a prolyl oligopeptidase domain. Moreover, other biologically active portions, in which other

regions of the protein are deleted, can be prepared by recombinant techniques and evaluated for one or more of the functional activities of a native 21953 protein.

[138] In a preferred embodiment, the 21953 protein has an amino acid sequence shown in SEQ ID NO:2. In other embodiments, the 21953 protein is substantially identical to SEQ ID NO:2. In yet another embodiment, the 21953 protein is substantially identical to SEQ ID NO:2 and retains the functional activity of the protein of SEQ ID NO:2, as described in detail in the subsections above.

[139] In another preferred embodiment, the 21953 protein has a  $K_m$  for the substrate H-Gly-Pro-*p*-nitroanilide (NA)/HCl (Sigma Corp, MO, USA) (H-Gly-Pro-pNA) of less than about 10 mM, 5 mM, 1 mM, 0.5 mM, 0.2 mM, or 0.1 mM, and/or a  $V_{max}$  for H-Gly-Pro-pNA of about at least 100, 500, 1000, 3000, 5000, or 10000 absorbance units·min<sup>-1</sup>. Such parameters can be determined using a prolyl oligopeptidase assay described herein, e.g., as described in "Screening Assays," below.

### **21953 Chimeric or Fusion Proteins**

[140] In another aspect, the invention provides 21953 chimeric or fusion proteins. As used herein, a 21953 "chimeric protein" or "fusion protein" includes a 21953 polypeptide linked to a non-21953 polypeptide. A "non-21953 polypeptide" refers to a polypeptide having an amino acid sequence corresponding to a protein which is not substantially homologous to the 21953 protein, e.g., a protein which is different from the 21953 protein and which is derived from the same or a different organism. The 21953 polypeptide of the fusion protein can correspond to all or a portion e.g., a fragment described herein of a 21953 amino acid sequence. In a preferred embodiment, a 21953 fusion protein includes at least one (or two) biologically active portion of a 21953 protein. The non-21953 polypeptide can be fused to the N-terminus or C-terminus of the 21953 polypeptide.

[141] The fusion protein can include a moiety which has a high affinity for a ligand. For example, the fusion protein can be a GST-21953 fusion protein in which the 21953 sequences are fused to the C-terminus of the GST sequences. Such fusion proteins can facilitate the purification of recombinant 21953. Alternatively, the fusion protein can be a 21953 protein containing a heterologous signal sequence at its N-terminus. In certain host cells (e.g., mammalian host cells), expression and/or secretion of 21953 can be increased through use of a heterologous signal sequence.

[142] Fusion proteins can include all or a part of a serum protein, e.g., an IgG constant region, or human serum albumin.

[143] The 21953 fusion proteins of the invention can be incorporated into pharmaceutical compositions and administered to a subject *in vivo*. The 21953 fusion proteins can be used to affect the bioavailability of a 21953 substrate. 21953 fusion proteins may be useful therapeutically for the treatment of disorders caused by, for example, (i) aberrant modification or mutation of a gene encoding a 21953 protein; (ii) mis-regulation of the 21953 gene; and (iii) aberrant post-translational modification of a 21953 protein.

[144] Moreover, the 21953-fusion proteins of the invention can be used as immunogens to produce anti-21953 antibodies in a subject, to purify 21953 ligands and in screening assays to identify molecules which inhibit the interaction of 21953 with a 21953 substrate.

[145] Expression vectors are commercially available that already encode a fusion moiety (e.g., a GST polypeptide). A 21953-encoding nucleic acid can be cloned into such an expression vector such that the fusion moiety is linked in-frame to the 21953 protein.

#### Variants of 21953 Proteins

[146] In another aspect, the invention also features a variant of a 21953 polypeptide, e.g., which functions as an agonist (mimetics) or as an antagonist. Variants of the 21953 proteins can be generated by mutagenesis, e.g., discrete point mutation, the insertion or deletion of sequences or the truncation of a 21953 protein. An agonist of the 21953 proteins can retain substantially the same, or a subset, of the biological activities of the naturally occurring form of a 21953 protein. An antagonist of a 21953 protein can inhibit one or more of the activities of the naturally occurring form of the 21953 protein by, for example, competitively modulating a 21953-mediated activity of a 21953 protein. Thus, specific biological effects can be elicited by treatment with a variant of limited function. Preferably, treatment of a subject with a variant having a subset of the biological activities of the naturally occurring form of the protein has fewer side effects in a subject relative to treatment with the naturally occurring form of the 21953 protein.

[147] Variants of a 21953 protein can be identified by screening combinatorial libraries of mutants, e.g., truncation mutants, of a 21953 protein for agonist or antagonist activity.

[148] Libraries of fragments e.g., N terminal, C terminal, or internal fragments, of a 21953 protein coding sequence can be used to generate a variegated population of fragments for screening and subsequent selection of variants of a 21953 protein. Variants in which a cysteine residues is added or deleted or in which a residue which is glycosylated is added or deleted are particularly preferred.

[149] Methods for screening gene products of combinatorial libraries made by point mutations or truncation, and for screening cDNA libraries for gene products having a selected property are known in the art. Such methods are adaptable for rapid screening of the gene libraries generated by combinatorial mutagenesis of 21953 proteins. Recursive ensemble mutagenesis (REM), a new technique which enhances the frequency of functional mutants in the libraries, can be used in combination with the screening assays to identify 21953 variants (Arkin and Yourvan (1992) *Proc. Natl. Acad. Sci. USA* 89:7811-7815; Delgrave *et al.* (1993) *Protein Engineering* 6:327-331).

[150] Cell based assays can be exploited to analyze a variegated 21953 library. For example, a library of expression vectors can be transfected into a cell line, e.g., a cell line, which ordinarily responds to 21953 in a substrate-dependent manner. The transfected cells are then contacted with 21953 and the effect of the expression of the mutant on signaling by the 21953 substrate can be detected, e.g., by measuring prolyl oligopeptidase as described below. Plasmid DNA can then be recovered from the cells which score for inhibition, or alternatively, potentiation of signaling by the 21953 substrate, and the individual clones further characterized.

[151] In another aspect, the invention features a method of making a 21953 polypeptide, e.g., a peptide having a non-wild type activity, e.g., an antagonist, agonist, or super agonist of a naturally occurring 21953 polypeptide, e.g., a naturally occurring 21953 polypeptide. The method includes: altering the sequence of a 21953 polypeptide, e.g., altering the sequence , e.g., by substitution or deletion of one or more residues of a non-conserved region, a domain or residue disclosed herein, and testing the altered polypeptide for the desired activity.

[152] In another aspect, the invention features a method of making a fragment or analog of a 21953 polypeptide a biological activity of a naturally occurring 21953 polypeptide. The method includes: altering the sequence, e.g., by substitution or deletion of one or more residues, of a 21953 polypeptide, e.g., altering the sequence of a non-conserved

region, or a domain or residue described herein, and testing the altered polypeptide for the desired activity.

#### Anti-21953 Antibodies

[153] In another aspect, the invention provides an anti-21953 antibody, or a fragment thereof (e.g., an antigen-binding fragment thereof). The term "antibody" as used herein refers to an immunoglobulin molecule or immunologically active portion thereof, i.e., an antigen-binding portion. As used herein, the term "antibody" refers to a protein comprising at least one, and preferably two, heavy (H) chain variable regions (abbreviated herein as VH), and at least one and preferably two light (L) chain variable regions (abbreviated herein as VL). The VH and VL regions can be further subdivided into regions of hypervariability, termed "complementarity determining regions" ("CDR"), interspersed with regions that are more conserved, termed "framework regions" (FR). The extent of the framework region and CDR's has been precisely defined (see, Kabat, E.A., et al. (1991) *Sequences of Proteins of Immunological Interest, Fifth Edition*, U.S. Department of Health and Human Services, NIH Publication No. 91-3242, and Chothia, C. et al. (1987) J. Mol. Biol. 196:901-917, which are incorporated herein by reference). Each VH and VL is composed of three CDR's and four FRs, arranged from amino-terminus to carboxy-terminus in the following order: FR1, CDR1, FR2, CDR2, FR3, CDR3, FR4.

[154] The anti-21953 antibody can further include a heavy and light chain constant region, to thereby form a heavy and light immunoglobulin chain, respectively. In one embodiment, the antibody is a tetramer of two heavy immunoglobulin chains and two light immunoglobulin chains, wherein the heavy and light immunoglobulin chains are interconnected by, e.g., disulfide bonds. The heavy chain constant region is comprised of three domains, CH1, CH2 and CH3. The light chain constant region is comprised of one domain, CL. The variable region of the heavy and light chains contains a binding domain that interacts with an antigen. The constant regions of the antibodies typically mediate the binding of the antibody to host tissues or factors, including various cells of the immune system (e.g., effector cells) and the first component (Clq) of the classical complement system.

[155] As used herein, the term "immunoglobulin" refers to a protein consisting of one or more polypeptides substantially encoded by immunoglobulin genes. The recognized human immunoglobulin genes include the kappa, lambda, alpha (IgA1 and IgA2), gamma

(IgG1, IgG2, IgG3, IgG4), delta, epsilon and mu constant region genes, as well as the myriad immunoglobulin variable region genes. Full-length immunoglobulin "light chains" (about 25 KDa or 214 amino acids) are encoded by a variable region gene at the NH<sub>2</sub>-terminus (about 110 amino acids) and a kappa or lambda constant region gene at the COOH-terminus. Full-length immunoglobulin "heavy chains" (about 50 KDa or 446 amino acids), are similarly encoded by a variable region gene (about 116 amino acids) and one of the other aforementioned constant region genes, e.g., gamma (encoding about 330 amino acids).

[156] The term "antigen-binding fragment" of an antibody (or simply "antibody portion," or "fragment"), as used herein, refers to one or more fragments of a full-length antibody that retain the ability to specifically bind to the antigen, e.g., 21953 polypeptide or fragment thereof. Examples of antigen-binding fragments of the anti-21953 antibody include, but are not limited to: (i) a Fab fragment, a monovalent fragment consisting of the VL, VH, CL and CH1 domains; (ii) a F(ab')<sub>2</sub> fragment, a bivalent fragment comprising two Fab fragments linked by a disulfide bridge at the hinge region; (iii) a Fd fragment consisting of the VH and CH1 domains; (iv) a Fv fragment consisting of the VL and VH domains of a single arm of an antibody, (v) a dAb fragment (Ward *et al.*, (1989) *Nature* 341:544-546), which consists of a VH domain; and (vi) an isolated complementarity determining region (CDR). Furthermore, although the two domains of the Fv fragment, VL and VH, are coded for by separate genes, they can be joined, using recombinant methods, by a synthetic linker that enables them to be made as a single protein chain in which the VL and VH regions pair to form monovalent molecules (known as single chain Fv (scFv); see e.g., Bird *et al.* (1988) *Science* 242:423-426; and Huston *et al.* (1988) *Proc. Natl. Acad. Sci. USA* 85:5879-5883). Such single chain antibodies are also encompassed within the term "antigen-binding fragment" of an antibody. These antibody fragments are obtained using conventional techniques known to those with skill in the art, and the fragments are screened for utility in the same manner as are intact antibodies.

[157] The anti-21953 antibody can be a polyclonal or a monoclonal antibody. In other embodiments, the antibody can be recombinantly produced, e.g., produced by phage display or by combinatorial methods.

[158] Phage display and combinatorial methods for generating anti-21953 antibodies are known in the art (as described in, e.g., Ladner *et al.* U.S. Patent No. 5,223,409; Kang *et al.* International Publication No. WO 92/18619; Dower *et al.* International Publication No.

WO 91/17271; Winter et al. International Publication WO 92/20791; Markland et al. International Publication No. WO 92/15679; Breitling et al. International Publication WO 93/01288; McCafferty et al. International Publication No. WO 92/01047; Garrard et al. International Publication No. WO 92/09690; Ladner et al. International Publication No. WO 90/02809; Fuchs et al. (1991) *Bio/Technology* 9:1370-1372; Hay et al. (1992) *Hum Antibod Hybridomas* 3:81-85; Huse et al. (1989) *Science* 246:1275-1281; Griffiths et al. (1993) *EMBO J* 12:725-734; Hawkins et al. (1992) *J Mol Biol* 226:889-896; Clackson et al. (1991) *Nature* 352:624-628; Gram et al. (1992) *PNAS* 89:3576-3580; Garrad et al. (1991) *Bio/Technology* 9:1373-1377; Hoogenboom et al. (1991) *Nuc Acid Res* 19:4133-4137; and Barbas et al. (1991) *PNAS* 88:7978-7982, the contents of all of which are incorporated by reference herein).

[159] In one embodiment, the anti-21953 antibody is a fully human antibody (e.g., an antibody made in a mouse which has been genetically engineered to produce an antibody from a human immunoglobulin sequence), or a non-human antibody, e.g., a rodent (mouse or rat), goat, primate (e.g., monkey), camel antibody. Preferably, the non-human antibody is a rodent (mouse or rat antibody). Method of producing rodent antibodies are known in the art.

[160] Human monoclonal antibodies can be generated using transgenic mice carrying the human immunoglobulin genes rather than the mouse system. Splenocytes from these transgenic mice immunized with the antigen of interest are used to produce hybridomas that secrete human mAbs with specific affinities for epitopes from a human protein (see, e.g., Wood et al. International Application WO 91/00906, Kucherlapati et al. PCT publication WO 91/10741; Lonberg et al. International Application WO 92/03918; Kay et al. International Application 92/03917; Lonberg, N. et al. 1994 *Nature* 368:856-859; Green, L.L. et al. 1994 *Nature Genet.* 7:13-21; Morrison, S.L. et al. 1994 *Proc. Natl. Acad. Sci. USA* 81:6851-6855; Bruggeman et al. 1993 *Year Immunol.* 7:33-40; Tuailon et al. 1993 *PNAS* 90:3720-3724; Bruggeman et al. 1991 *Eur J Immunol.* 21:1323-1326).

[161] An anti-21953 antibody can be one in which the variable region, or a portion thereof, e.g., the CDR's, are generated in a non-human organism, e.g., a rat or mouse. Chimeric, CDR-grafted, and humanized antibodies are within the invention. Antibodies generated in a non-human organism, e.g., a rat or mouse, and then modified, e.g., in the variable framework or constant region, to decrease antigenicity in a human are within the invention.

[162] Chimeric antibodies can be produced by recombinant DNA techniques known in the art. For example, a gene encoding the Fc constant region of a murine (or other species) monoclonal antibody molecule is digested with restriction enzymes to remove the region encoding the murine Fc, and the equivalent portion of a gene encoding a human Fc constant region is substituted (see Robinson et al., International Patent Publication PCT/US86/02269; Akira, et al., European Patent Application 184,187; Taniguchi, M., European Patent Application 171,496; Morrison et al., European Patent Application 173,494; Neuberger et al., International Application WO 86/01533; Cabilly et al. U.S. Patent No. 4,816,567; Cabilly et al., European Patent Application 125,023; Better et al. (1988) *Science* 240:1041-1043); Liu et al. (1987) *PNAS* 84:3439-3443; Liu et al., 1987, *J. Immunol.* 139:3521-3526; Sun et al. (1987) *PNAS* 84:214-218; Nishimura et al., 1987, *Canc. Res.* 47:999-1005; Wood et al. (1985) *Nature* 314:446-449; and Shaw et al., 1988, *J. Natl Cancer Inst.* 80:1553-1559).

[163] A humanized or CDR-grafted antibody will have at least one or two but generally all three recipient CDR's (of heavy and or light immunoglobulin chains) replaced with a donor CDR. The antibody may be replaced with at least a portion of a non-human CDR or only some of the CDR's may be replaced with non-human CDR's. It is only necessary to replace the number of CDR's required for binding of the humanized antibody to a 21953 or a fragment thereof. Preferably, the donor will be a rodent antibody, e.g., a rat or mouse antibody, and the recipient will be a human framework or a human consensus framework. Typically, the immunoglobulin providing the CDR's is called the "donor" and the immunoglobulin providing the framework is called the "acceptor." In one embodiment, the donor immunoglobulin is a non-human (e.g., rodent). The acceptor framework is a naturally-occurring (e.g., a human) framework or a consensus framework, or a sequence about 85% or higher, preferably 90%, 95%, 99% or higher identical thereto.

[164] As used herein, the term "consensus sequence" refers to the sequence formed from the most frequently occurring amino acids (or nucleotides) in a family of related sequences (See e.g., Winnacker, From Genes to Clones (Verlagsgesellschaft, Weinheim, Germany 1987). In a family of proteins, each position in the consensus sequence is occupied by the amino acid occurring most frequently at that position in the family. If two amino acids occur equally frequently, either can be included in the consensus sequence. A "consensus framework" refers to the framework region in the consensus immunoglobulin sequence.

[165] An antibody can be humanized by methods known in the art. Humanized antibodies can be generated by replacing sequences of the Fv variable region which are not directly involved in antigen binding with equivalent sequences from human Fv variable regions. General methods for generating humanized antibodies are provided by Morrison, S. L., 1985, *Science* 229:1202-1207, by Oi et al., 1986, *BioTechniques* 4:214, and by Queen et al. US 5,585,089, US 5,693,761 and US 5,693,762, the contents of all of which are hereby incorporated by reference. Those methods include isolating, manipulating, and expressing the nucleic acid sequences that encode all or part of immunoglobulin Fv variable regions from at least one of a heavy or light chain. Sources of such nucleic acid are well known to those skilled in the art and, for example, may be obtained from a hybridoma producing an antibody against a 21953 polypeptide or fragment thereof. The recombinant DNA encoding the humanized antibody, or fragment thereof, can then be cloned into an appropriate expression vector.

[166] Humanized or CDR-grafted antibodies can be produced by CDR-grafting or CDR substitution; wherein one, two, or all CDR's of an immunoglobulin chain can be replaced. See, e.g., U.S. Patent 5,225,539; Jones et al. (1986) *Nature* 321:552-525; Verhoeyan et al. (1988) *Science* 239:1534; Beidler et al. (1988) *J. Immunol.* 141:4053-4060; Winter US 5,225,539, the contents of all of which are hereby expressly incorporated by reference. Winter describes a CDR-grafting method which may be used to prepare the humanized antibodies of the present invention (UK Patent Application GB 2188638A, filed on March 26, 1987; Winter US 5,225,539), the contents of which is expressly incorporated by reference.

[167] Also within the scope of the invention are humanized antibodies in which specific amino acids have been substituted, deleted or added. Preferred humanized antibodies have amino acid substitutions in the framework region, such as to improve binding to the antigen. For example, a humanized antibody will have framework residues identical to the donor framework residue or to another amino acid other than the recipient framework residue. To generate such antibodies, a selected, small number of acceptor framework residues of the humanized immunoglobulin chain can be replaced by the corresponding donor amino acids. Preferred locations of the substitutions include amino acid residues adjacent to the CDR, or which are capable of interacting with a CDR (see e.g., US 5,585,089). Criteria for selecting amino acids from the donor are described in US 5,585,089, e.g., columns 12-16 of US 5,585,089, the e.g., columns 12-16 of US 5,585,089,

the contents of which are hereby incorporated by reference. Other techniques for humanizing antibodies are described in Padlan et al. EP 519596 A1, published on December 23, 1992.

[168] In preferred embodiments an antibody can be made by immunizing with purified 21953 antigen, or a fragment thereof, e.g., a fragment described herein, tissue, e.g., crude tissue preparations, whole cells, preferably living cells, lysed cells, or cell fractions.

[169] A full-length 21953 protein or, antigenic peptide fragment of 21953 can be used as an immunogen or can be used to identify anti-21953 antibodies made with other immunogens, e.g., cells, membrane preparations, and the like. The antigenic peptide of 21953 should include at least 8 amino acid residues of the amino acid sequence shown in SEQ ID NO:2 and encompasses an epitope of 21953. Preferably, the antigenic peptide includes at least 10 amino acid residues, more preferably at least 15 amino acid residues, even more preferably at least 20 amino acid residues, and most preferably at least 30 amino acid residues.

[170] Hydrophilic fragments of 21953, e.g., those which include residues 20 to 40, 65 to 80, or 780 to 790, can be used to make, e.g., used as immunogens or used to characterize the specificity of an antibody, antibodies against hydrophilic regions of the 21953 protein. Similarly, a hydrophobic fragment of 21953, e.g. which include residues 250 to 270, 370 to 390, or 681 to 695, can be used to make an antibody against a hydrophobic region of the 21953 protein; a fragment of 21953 which include residues about 672 to 744, 672 to 690, 690 to 710, or 710 to 744 can be used to make an antibody against the prolyl oligopeptidase domain of the 21953 protein.

[171] Antibodies reactive with, or specific for, any of these regions, or other regions or domains described herein are provided.

[172] Antibodies which bind only native 21953 protein, only denatured or otherwise non-native 21953 protein, or which bind both, are within the invention. Antibodies with linear or conformational epitopes are within the invention. Conformational epitopes can sometimes be identified by identifying antibodies which bind to native but not denatured 21953 protein.

[173] Preferred epitopes encompassed by the antigenic peptide are regions of 21953 are located on the surface of the protein, e.g., hydrophilic regions, as well as regions with high antigenicity. For example, an Emini surface probability analysis of the human 21953 protein sequence can be used to indicate the regions that have a particularly high probability

of being localized to the surface of the 21953 protein and are thus likely to constitute surface residues useful for targeting antibody production.

[174] The anti-21953 antibody can be a single chain antibody. A single-chain antibody (scFV) may be engineered (see, for example, Colcher, D. *et al.* (1999) *Ann NY Acad Sci* 880:263-80; and Reiter, Y. (1996) *Clin Cancer Res* 2:245-52). The single chain antibody can be dimerized or multimerized to generate multivalent antibodies having specificities for different epitopes of the same target 21953 protein.

[175] In a preferred embodiment the antibody has: effector function; and can fix complement. In other embodiments the antibody does not; recruit effector cells; or fix complement.

[176] In a preferred embodiment, the antibody has reduced or no ability to bind an Fc receptor. For example., it is a isotype or subtype, fragment or other mutant, which does not support binding to an Fc receptor, e.g., it has a mutagenized or deleted Fc receptor binding region.

[177] In a preferred embodiment, an anti-21953 antibody alters (e.g., increases or decreases) the prolyl oligopeptidase activity of a 21953 polypeptide. For example, the antibody can specifically bind a residue of the active site of 21953 polypeptide, e.g., a residue located between about 730 to 750, 805 to 830, 835 to 860 of SEQ ID NO:2. The antibody can block the binding of substrate to the 21953 polypeptide.

[178] In another preferred embodiment, the antibody specifically binds a residue in the prolyl oligopeptidase domain, e.g., from about amino acid 672 to 744, or 610 to 883 of SEQ ID NO:2, or in the DPP IV N-terminal residue, e.g., a residue between about amino acids 88 to 663 of SEQ ID NO:2.

[179] The antibody can be coupled to a toxin, e.g., a polypeptide toxin, e,g, ricin or diphtheria toxin or active fragment hereof, or a radioactive nucleus, or imaging agent, e.g. a radioactive, enzymatic, or other, e.g., imaging agent, e.g., a NMR contrast agent. Labels which produce detectable radioactive emissions or fluorescence are preferred.

[180] An anti-21953 antibody (e.g., monoclonal antibody) can be used to isolate 21953 by standard techniques, such as affinity chromatography or immunoprecipitation. Moreover, an anti-21953 antibody can be used to detect 21953 protein (e.g., in a cellular lysate or cell supernatant) in order to evaluate the abundance and pattern of expression of the protein. Anti-21953 antibodies can be used diagnostically to monitor protein levels in tissue as part of a clinical testing procedure, e.g., to determine the efficacy of a given treatment

regimen. Detection can be facilitated by coupling (i.e., physically linking) the antibody to a detectable substance (i.e., antibody labelling). Examples of detectable substances include various enzymes, prosthetic groups, fluorescent materials, luminescent materials, bioluminescent materials, and radioactive materials. Examples of suitable enzymes include horseradish peroxidase, alkaline phosphatase,  $\beta$ -galactosidase, or acetylcholinesterase; examples of suitable prosthetic group complexes include streptavidin/biotin and avidin/biotin; examples of suitable fluorescent materials include umbelliflone, fluorescein, fluorescein isothiocyanate, rhodamine, dichlorotriazinylamine fluorescein, dansyl chloride or phycoerythrin; an example of a luminescent material includes luminol; examples of bioluminescent materials include luciferase, luciferin, and aequorin, and examples of suitable radioactive material include  $^{125}\text{I}$ ,  $^{131}\text{I}$ ,  $^{35}\text{S}$  or  $^3\text{H}$ .

[181] The invention also includes a nucleic acid which encodes an anti-21953 antibody, e.g., an anti-21953 antibody described herein. Also included are vectors which include the nucleic acid and cells transformed with the nucleic acid, particularly cells which are useful for producing an antibody, e.g., mammalian cells, e.g. CHO or lymphatic cells.

[182] The invention also includes cell lines, e.g., hybridomas, which make an anti-21953 antibody, e.g., and antibody described herein, and method of using said cells to make a 21953 antibody.

#### Recombinant Expression Vectors, Host Cells and Genetically Engineered Cells

[183] In another aspect, the invention includes, vectors, preferably expression vectors, containing a nucleic acid encoding a polypeptide described herein. As used herein, the term "vector" refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked and can include a plasmid, cosmid or viral vector. The vector can be capable of autonomous replication or it can integrate into a host DNA. Viral vectors include, e.g., replication defective retroviruses, adenoviruses and adeno-associated viruses.

[184] A vector can include a 21953 nucleic acid in a form suitable for expression of the nucleic acid in a host cell. Preferably the recombinant expression vector includes one or more regulatory sequences operatively linked to the nucleic acid sequence to be expressed. The term "regulatory sequence" includes promoters, enhancers and other expression control elements (e.g., polyadenylation signals). Regulatory sequences include those which direct constitutive expression of a nucleotide sequence, as well as tissue-specific regulatory and/or inducible sequences. The design of the expression vector can depend on such factors as the

choice of the host cell to be transformed, the level of expression of protein desired, and the like. The expression vectors of the invention can be introduced into host cells to thereby produce proteins or polypeptides, including fusion proteins or polypeptides, encoded by nucleic acids as described herein (e.g., 21953 proteins, mutant forms of 21953 proteins, fusion proteins, and the like).

[185] The recombinant expression vectors of the invention can be designed for expression of 21953 proteins in prokaryotic or eukaryotic cells. For example, polypeptides of the invention can be expressed in *E. coli*, insect cells (e.g., using baculovirus expression vectors), yeast cells or mammalian cells. Suitable host cells are discussed further in Goeddel, (1990) *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, CA. Alternatively, the recombinant expression vector can be transcribed and translated *in vitro*, for example using T7 promoter regulatory sequences and T7 polymerase.

[186] Expression of proteins in prokaryotes is most often carried out in *E. coli* with vectors containing constitutive or inducible promoters directing the expression of either fusion or non-fusion proteins. Fusion-vectors add a number of amino acids to a protein encoded therein, usually to the amino terminus of the recombinant protein. Such fusion vectors typically serve three purposes: 1) to increase expression of recombinant protein; 2) to increase the solubility of the recombinant protein; and 3) to aid in the purification of the recombinant protein by acting as a ligand in affinity purification. Often, a proteolytic cleavage site is introduced at the junction of the fusion moiety and the recombinant protein to enable separation of the recombinant protein from the fusion moiety subsequent to purification of the fusion protein. Such enzymes, and their cognate recognition sequences, include Factor Xa, thrombin and enterokinase. Typical fusion expression vectors include pGEX (Pharmacia Biotech Inc; Smith, D.B. and Johnson, K.S. (1988) *Gene* 67:31-40), pMAL (New England Biolabs, Beverly, MA) and pRIT5 (Pharmacia, Piscataway, NJ) which fuse glutathione S-transferase (GST), maltose E binding protein, or protein A, respectively, to the target recombinant protein. Additional convenient fusion moieties include the hexa-histidine tag which can be inserted in frame at either terminus of coding region, or in loop regions or inter-domain linkers. A polypeptide that includes a hexa-histidine tag can be purified by immobilized metal chelate chromatography, e.g., using Ni<sup>2+</sup>-NTA resin (Qiagen, Inc.).

[187] Purified fusion proteins can be used in 21953 activity assays, (e.g., direct assays or competitive assays described in detail below), or to generate antibodies specific for 21953 proteins. In a preferred embodiment, a fusion protein expressed in a retroviral expression vector of the present invention can be used to infect bone marrow cells which are subsequently transplanted into irradiated recipients. The pathology of the subject recipient is then examined after sufficient time has passed (e.g., six weeks).

[188] To maximize recombinant protein expression in *E. coli* is to express the protein in a host bacteria with an impaired capacity to proteolytically cleave the recombinant protein (Gottesman, S., (1990) *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, California 119-128). Another strategy is to alter the nucleic acid sequence of the nucleic acid to be inserted into an expression vector so that the individual codons for each amino acid are those preferentially utilized in *E. coli* (Wada *et al.*, (1992) *Nucleic Acids Res.* 20:2111-2118). Such alteration of nucleic acid sequences of the invention can be carried out by standard DNA synthesis techniques.

[189] The 21953 expression vector can be a yeast expression vector, a vector for expression in insect cells, e.g., a baculovirus expression vector or a vector suitable for expression in mammalian cells.

[190] When used in mammalian cells, the expression vector's control functions can be provided by viral regulatory elements. For example, commonly used promoters are derived from polyoma, Adenovirus 2, cytomegalovirus and Simian Virus 40.

[191] In another embodiment, the promoter is an inducible promoter, e.g., a promoter regulated by a steroid hormone, by a polypeptide hormone (e.g., by means of a signal transduction pathway), or by a heterologous polypeptide (e.g., the tetracycline-inducible systems, "Tet-On" and "Tet-Off"; see, e.g., Clontech Inc., CA, Gossen and Bujard (1992) *Proc. Natl. Acad. Sci. USA* 89:5547, and Paillard (1989) *Human Gene Therapy* 9:983).

[192] In another embodiment, the recombinant mammalian expression vector is capable of directing expression of the nucleic acid preferentially in a particular cell type (e.g., tissue-specific regulatory elements are used to express the nucleic acid). Non-limiting examples of suitable tissue-specific promoters include the albumin promoter (liver-specific; Pinkert *et al.* (1987) *Genes Dev.* 1:268-277), lymphoid-specific promoters (Calame and Eaton (1988) *Adv. Immunol.* 43:235-275), in particular promoters of T cell receptors (Winoto and Baltimore (1989) *EMBO J.* 8:729-733) and immunoglobulins (Banerji *et al.* (1983) *Cell* 33:729-740; Queen and Baltimore (1983) *Cell* 33:741-748), neuron-specific

promoters (e.g., the neurofilament promoter; Byrne and Ruddle (1989) *Proc. Natl. Acad. Sci. USA* 86:5473-5477), pancreas-specific promoters (Edlund *et al.* (1985) *Science* 230:912-916), and mammary gland-specific promoters (e.g., milk whey promoter; U.S. Patent No. 4,873,316 and European Application Publication No. 264,166).

Developmentally-regulated promoters are also encompassed, for example, the murine hox promoters (Kessel and Gruss (1990) *Science* 249:374-379) and the  $\alpha$ -fetoprotein promoter (Campes and Tilghman (1989) *Genes Dev.* 3:537-546).

[193] The invention further provides a recombinant expression vector comprising a DNA molecule of the invention cloned into the expression vector in an antisense orientation. Regulatory sequences (e.g., viral promoters and/or enhancers) operatively linked to a nucleic acid cloned in the antisense orientation can be chosen which direct the constitutive, tissue specific or cell type specific expression of antisense RNA in a variety of cell types. The antisense expression vector can be in the form of a recombinant plasmid, phagemid or attenuated virus.

[194] Another aspect the invention provides a host cell which includes a nucleic acid molecule described herein; e.g., a 21953 nucleic acid molecule within a recombinant expression vector or a 21953 nucleic acid molecule containing sequences which allow it to homologously recombine into a specific site of the host cell's genome. The terms "host cell" and "recombinant host cell" are used interchangeably herein. Such terms refer not only to the particular subject cell but to the progeny or potential progeny of such a cell. Because certain modifications may occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term as used herein.

[195] A host cell can be any prokaryotic or eukaryotic cell. For example, a 21953 protein can be expressed in bacterial cells (such as *E. coli*), insect cells, yeast or mammalian cells (such as Chinese hamster ovary cells (CHO) or COS cells). Other suitable host cells are known to those skilled in the art.

[196] Vector DNA can be introduced into host cells via conventional transformation or transfection techniques. As used herein, the terms "transformation" and "transfection" are intended to refer to a variety of art-recognized techniques for introducing foreign nucleic acid (e.g., DNA) into a host cell, including calcium phosphate or calcium chloride co-precipitation, DEAE-dextran-mediated transfection, lipofection, or electroporation.

[197] A host cell of the invention can be used to produce (i.e., express) a 21953 protein. Accordingly, the invention further provides methods for producing a 21953 protein using the host cells of the invention. In one embodiment, the method includes culturing the host cell of the invention (into which a recombinant expression vector encoding a 21953 protein has been introduced) in a suitable medium such that a 21953 protein is produced. In another embodiment, the method further includes isolating a 21953 protein from the medium or the host cell.

[198] In another aspect, the invention features, a cell or purified preparation of cells which include a 21953 transgene, or which otherwise misexpress 21953. The cell preparation can consist of human or non-human cells, e.g., rodent cells, e.g., mouse or rat cells, rabbit cells, or pig cells. In preferred embodiments, the cell or cells include a 21953 transgene, e.g., a heterologous form of a 21953, e.g., a gene derived from humans (in the case of a non-human cell). The 21953 transgene can be misexpressed, e.g., overexpressed or underexpressed. In other preferred embodiments, the cell or cells include a gene that misexpresses an endogenous 21953, e.g., a gene the expression of which is disrupted, e.g., a knockout. Such cells can serve as a model for studying disorders that are related to mutated or mis-expressed 21953 alleles or for use in drug screening.

[199] In another aspect, the invention features, a human cell, e.g., a hematopoietic stem cell, transformed with nucleic acid which encodes a subject 21953 polypeptide.

[200] Also provided are cells, preferably human cells, e.g., human hematopoietic or fibroblast cells, in which an endogenous 21953 is under the control of a regulatory sequence that does not normally control the expression of the endogenous 21953 gene. The expression characteristics of an endogenous gene within a cell, e.g., a cell line or microorganism, can be modified by inserting a heterologous DNA regulatory element into the genome of the cell such that the inserted regulatory element is operably linked to the endogenous 21953 gene. For example, an endogenous 21953 gene which is "transcriptionally silent," e.g., not normally expressed, or expressed only at very low levels, may be activated by inserting a regulatory element which is capable of promoting the expression of a normally expressed gene product in that cell. Techniques such as targeted homologous recombinations, can be used to insert the heterologous DNA as described in, e.g., Chappel, US 5,272,071; WO 91/06667, published in May 16, 1991.

[201] In a preferred embodiment, recombinant cells described herein can be used for replacement therapy in a subject. For example, a nucleic acid encoding a 21953 polypeptide

operably linked to an inducible promoter (e.g., a steroid hormone receptor-regulated promoter) is introduced into a human or nonhuman, e.g., mammalian, e.g., porcine recombinant cell. The cell is cultivated and encapsulated in a biocompatible material, such as poly-lysine alginate, and subsequently implanted into the subject. See, e.g., Lanza (1996) *Nat. Biotechnol.* 14:1107; Joki *et al.* (2001) *Nat. Biotechnol.* 19:35; and U.S. Patent No. 5,876,742. Production of 21953 polypeptide can be regulated in the subject by administering an agent (e.g., a steroid hormone) to the subject. In another preferred embodiment, the implanted recombinant cells express and secrete an antibody specific for a 21953 polypeptide. The antibody can be any antibody or any antibody derivative described herein.

#### Transgenic Animals

[202] The invention provides non-human transgenic animals. Such animals are useful for studying the function and/or activity of a 21953 protein and for identifying and/or evaluating modulators of 21953 activity. As used herein, a "transgenic animal" is a non-human animal, preferably a mammal, more preferably a rodent such as a rat or mouse, in which one or more of the cells of the animal includes a transgene. Other examples of transgenic animals include non-human primates, sheep, dogs, cows, goats, chickens, amphibians, and the like. A transgene is exogenous DNA or a rearrangement, e.g., a deletion of endogenous chromosomal DNA, which preferably is integrated into or occurs in the genome of the cells of a transgenic animal. A transgene can direct the expression of an encoded gene product in one or more cell types or tissues of the transgenic animal, other transgenes, e.g., a knockout, reduce expression. Thus, a transgenic animal can be one in which an endogenous 21953 gene has been altered by, e.g., by homologous recombination between the endogenous gene and an exogenous DNA molecule introduced into a cell of the animal, e.g., an embryonic cell of the animal, prior to development of the animal.

[203] Intronic sequences and polyadenylation signals can also be included in the transgene to increase the efficiency of expression of the transgene. A tissue-specific regulatory sequence(s) can be operably linked to a transgene of the invention to direct expression of a 21953 protein to particular cells. A transgenic founder animal can be identified based upon the presence of a 21953 transgene in its genome and/or expression of 21953 mRNA in tissues or cells of the animals. A transgenic founder animal can then be used to breed additional animals carrying the transgene. Moreover, transgenic animals

carrying a transgene encoding a 21953 protein can further be bred to other transgenic animals carrying other transgenes.

[204] 21953 proteins or polypeptides can be expressed in transgenic animals or plants, e.g., a nucleic acid encoding the protein or polypeptide can be introduced into the genome of an animal. In preferred embodiments the nucleic acid is placed under the control of a tissue specific promoter, e.g., a milk or egg specific promoter, and recovered from the milk or eggs produced by the animal. Suitable animals are mice, pigs, cows, goats, and sheep.

[205] The invention also includes a population of cells from a transgenic animal, as discussed, e.g., below.

#### Uses

[206] The nucleic acid molecules, proteins, protein homologues, and antibodies described herein can be used in one or more of the following methods: a) screening assays; b) predictive medicine (e.g., diagnostic assays, prognostic assays, monitoring clinical trials, and pharmacogenetics); c) methods of treatment (e.g., therapeutic and prophylactic); and d) *in vitro* modification of polypeptide hormones.

[207] The isolated nucleic acid molecules of the invention can be used, for example, to express a 21953 protein (e.g., via a recombinant expression vector in a host cell in gene therapy applications), to detect a 21953 mRNA (e.g., in a biological sample) or a genetic alteration in a 21953 gene, and to modulate 21953 activity, as described further below. The 21953 proteins can be used to treat disorders characterized by insufficient or excessive production of a 21953 substrate or production of 21953 inhibitors. In addition, the 21953 proteins can be used to screen for naturally occurring 21953 substrates, to screen for drugs or compounds which modulate 21953 activity, as well as to treat disorders characterized by insufficient or excessive production of 21953 protein or production of 21953 protein forms which have decreased, aberrant or unwanted activity compared to 21953 wild type protein (e.g., a cell proliferative or cell differentiative disorder, e.g., a cancer, e.g., a cancer of the lung, prostate, breast, ovary, or colon). Moreover, the anti-21953 antibodies of the invention can be used to detect and isolate 21953 proteins, regulate the bioavailability of 21953 proteins, and modulate 21953 activity.

[208] A method of evaluating a compound for the ability to interact with, e.g., bind, a subject 21953 polypeptide is provided. The method includes: contacting the compound with the subject 21953 polypeptide; and evaluating ability of the compound to interact with,

4

e.g., to bind or form a complex with the subject 21953 polypeptide. This method can be performed *in vitro*, e.g., in a cell free system, or *in vivo*, e.g., in a two-hybrid interaction trap assay. This method can be used to identify naturally occurring molecules that interact with subject 21953 polypeptide. It can also be used to find natural or synthetic inhibitors of subject 21953 polypeptide. Screening methods are discussed in more detail below.

[209] The 21953 polypeptide is also an enzyme useful for processing polypeptide hormone precursors. For example, the 21953 polypeptide can be used in a method that includes a) providing a polypeptide hormone precursor; b) combining the polypeptide hormone polypeptide with a 21953 polypeptide or active fragment thereof (e.g., in an effective amount) to provide a reaction mixture; and c) maintaining the mixture under conditions such that the polypeptide hormone precursor is modified to yield the processed polypeptide hormone, e.g., an active form thereof. The method can further include d) separating the processed polypeptide hormone from the 21953 polypeptide. The polypeptide hormone precursor can be obtained from a synthetic process or from a producing cell. The method can be used in the preparation of a pharmaceutical composition that includes the processed hormone.

### Screening Assays

[210] The invention provides methods (also referred to herein as "screening assays") for identifying modulators, i.e., candidate or test compounds or agents (e.g., proteins, peptides, peptidomimetics, peptoids, small molecules or other drugs) which bind to 21953 proteins, have a stimulatory or inhibitory effect on, for example, 21953 expression or 21953 activity, or have a stimulatory or inhibitory effect on, for example, the expression or activity of a 21953 substrate. Compounds thus identified can be used to modulate the activity of target gene products (e.g., 21953 genes) in a therapeutic protocol, to elaborate the biological function of the target gene product, or to identify compounds that disrupt normal target gene interactions.

[211] The prolyl oligopeptidase activity of a 21953 polypeptide can be assayed *in vitro* using an enzymatic assay such as described in Abbott *et al.* (1999) *FEBS Lett.* 458:278-284 and Abbott *et al.* (2000) *Eur. J. Biochem* 267:6140-4150. A sample to be assayed is combined with substrate in phosphate buffer pH 7.4. Substrates include H-Gly-Pro-*p*-nitroanilide (NA)/HCl (Sigma Corp, MO, USA), and Gly-Pro-7-amino-4-trifluoromethylcoumarin (Calbiochem, San Diego, CA, USA) and other peptidyl substrates.

The reaction is incubated for 30 minutes at 37°C. For example, hydrolysis of H-Gly-Pro-pNA is monitored spectroscopically at 405 nm. The sample to be assayed can be a purified 21953 polypeptide, e.g., a 21953 polypeptide or a 21953 fusion protein purified by a method described herein. Routine Michaelis-Menten analysis of kinetic parameters can be used to quantify the enzymatic activity. Alternatively, the reaction can be quenched and total substrate hydrolyzed can be measured as indication of the activity.

[212] In one embodiment, the invention provides assays for screening candidate or test compounds which are substrates of a 21953 protein or polypeptide or a biologically active portion thereof. In another embodiment, the invention provides assays for screening candidate or test compounds that bind to or modulate the activity of a 21953 protein or polypeptide or a biologically active portion thereof. The afore-mentioned assay can be used by adding a candidate or test compound to the reaction mixture, either before or together with addition of the substrate.

[213] The test compounds of the present invention can be obtained using any of the numerous approaches in combinatorial library methods known in the art, including: biological libraries; peptoid libraries (libraries of molecules having the functionalities of peptides, but with a novel, non-peptide backbone which are resistant to enzymatic degradation but which nevertheless remain bioactive; see, e.g., Zuckermann, R.N. *et al.* (1994) *J. Med. Chem.* 37:2678-85); spatially addressable parallel solid phase or solution phase libraries; synthetic library methods requiring deconvolution; the 'one-bead one-compound' library method; and synthetic library methods using affinity chromatography selection. The biological library and peptoid library approaches are limited to peptide libraries, while the other four approaches are applicable to peptide, non-peptide oligomer or small molecule libraries of compounds (Lam (1997) *Anticancer Drug Des.* 12:145).

[214] Examples of methods for the synthesis of molecular libraries can be found in the art, for example in: DeWitt *et al.* (1993) *Proc. Natl. Acad. Sci. U.S.A.* 90:6909; Erb *et al.* (1994) *Proc. Natl. Acad. Sci. USA* 91:11422; Zuckermann *et al.* (1994). *J. Med. Chem.* 37:2678; Cho *et al.* (1993) *Science* 261:1303; Carell *et al.* (1994) *Angew. Chem. Int. Ed. Engl.* 33:2059; Carell *et al.* (1994) *Angew. Chem. Int. Ed. Engl.* 33:2061; and Gallop *et al.* (1994) *J. Med. Chem.* 37:1233. Combinatorial chemical libraries can be designed based on known substrates of oligopeptidases. For example, compounds can be designed that are peptidomimetics, e.g., phosphonate analogs of a peptide substrate, such as a proline-containing peptide.

[215] Libraries of compounds may be presented in solution (e.g., Houghten (1992) *Biotechniques* 13:412-421), or on beads (Lam (1991) *Nature* 354:82-84), chips (Fodor (1993) *Nature* 364:555-556), bacteria (Ladner, U.S. Patent No. 5,223,409), spores (Ladner U.S. Patent No. 5,223,409), plasmids (Cull *et al.* (1992) *Proc Natl Acad Sci USA* 89:1865-1869) or on phage (Scott and Smith (1990) *Science* 249:386-390; Devlin (1990) *Science* 249:404-406; Cwirla *et al.* (1990) *Proc. Natl. Acad. Sci.* 87:6378-6382; Felici (1991) *J. Mol. Biol.* 222:301-310; Ladner *supra*).

[216] In one embodiment, an assay is a cell-based assay in which a cell which expresses a 21953 protein or biologically active portion thereof is contacted with a test compound, and the ability of the test compound to modulate 21953 activity is determined. Determining the ability of the test compound to modulate 21953 activity can be accomplished by monitoring, for example, prolyl oligopeptidase activity. The cell, for example, can be of mammalian origin, e.g., human.

[217] The ability of the test compound to modulate 21953 binding to a compound, e.g., a 21953 substrate, or to bind to 21953 can also be evaluated. This can be accomplished, for example, by coupling the compound, e.g., the substrate, with a radioisotope or enzymatic label such that binding of the compound, e.g., the substrate, to 21953 can be determined by detecting the labeled compound, e.g., substrate, in a complex. Alternatively, 21953 could be coupled with a radioisotope or enzymatic label to monitor the ability of a test compound to modulate 21953 binding to a 21953 substrate in a complex. For example, compounds (e.g., 21953 substrates) can be labeled with  $^{125}\text{I}$ ,  $^{35}\text{S}$ ,  $^{14}\text{C}$ , or  $^3\text{H}$ , either directly or indirectly, and the radioisotope detected by direct counting of radioemission or by scintillation counting. Alternatively, compounds can be enzymatically labeled with, for example, horseradish peroxidase, alkaline phosphatase, or luciferase, and the enzymatic label detected by determination of conversion of an appropriate substrate to product.

[218] The ability of a compound (e.g., a 21953 substrate) to interact with 21953 with or without the labeling of any of the interactants can be evaluated. For example, a microphysiometer can be used to detect the interaction of a compound with 21953 without the labeling of either the compound or the 21953. McConnell, H. M. *et al.* (1992) *Science* 257:1906-1912. As used herein, a "microphysiometer" (e.g., CytoSensor) is an analytical instrument that measures the rate at which a cell acidifies its environment using a light-

addressable potentiometric sensor (LAPS). Changes in this acidification rate can be used as an indicator of the interaction between a compound and 21953.

[219] In yet another embodiment, a cell-free assay is provided in which a 21953 protein or biologically active portion thereof is contacted with a test compound and the ability of the test compound to bind to the 21953 protein or biologically active portion thereof is evaluated. Preferred biologically active portions of the 21953 proteins to be used in assays of the present invention include fragments which participate in interactions with non-21953 molecules, e.g., fragments with high surface probability scores.

[220] Soluble and/or membrane-bound forms of isolated proteins (e.g., 21953 proteins or biologically active portions thereof) can be used in the cell-free assays of the invention. When membrane-bound forms of the protein are used, it may be desirable to utilize a solubilizing agent. Examples of such solubilizing agents include non-ionic detergents such as n-octylglucoside, n-dodecylglucoside, n-dodecylmaltoside, octanoyl-N-methylglucamide, decanoyl-N-methylglucamide, Triton® X-100, Triton® X-114, Thesit®, Isotridecypoly(ethylene glycol ether)<sub>n</sub>, 3-[*(3-cholamidopropyl)dimethylammonio*]-1-propane sulfonate (CHAPS), 3-[*(3-cholamidopropyl)dimethylammonio*]-2-hydroxy-1-propane sulfonate (CHAPSO), or N-dodecyl=N,N-dimethyl-3-ammonio-1-propane sulfonate.

[221] Cell-free assays involve preparing a reaction mixture of the target gene protein and the test compound under conditions and for a time sufficient to allow the two components to interact and bind, thus forming a complex that can be removed and/or detected.

[222] The interaction between two molecules can also be detected, e.g., using fluorescence energy transfer (FET) (see, for example, Lakowicz *et al.*, U.S. Patent No. 5,631,169; Stavrianopoulos, *et al.*, U.S. Patent No. 4,868,103). A fluorophore label on the first, 'donor' molecule is selected such that its emitted fluorescent energy will be absorbed by a fluorescent label on a second, 'acceptor' molecule, which in turn is able to fluoresce due to the absorbed energy. Alternately, the 'donor' protein molecule may simply utilize the natural fluorescent energy of tryptophan residues. Labels are chosen that emit different wavelengths of light, such that the 'acceptor' molecule label may be differentiated from that of the 'donor'. Since the efficiency of energy transfer between the labels is related to the distance separating the molecules, the spatial relationship between the molecules can be assessed. In a situation in which binding occurs between the molecules, the fluorescent emission of the 'acceptor' molecule label in the assay should be maximal. An FET binding

event can be conveniently measured through standard fluorometric detection means well known in the art (e.g., using a fluorimeter).

[223] In another embodiment, determining the ability of the 21953 protein to bind to a target molecule can be accomplished using real-time Biomolecular Interaction Analysis (BIA) (see, e.g., Sjolander, S. and Urbaniczky, C. (1991) *Anal. Chem.* 63:2338-2345 and Szabo *et al.* (1995) *Curr. Opin. Struct. Biol.* 5:699-705). "Surface plasmon resonance" or "BIA" detects biospecific interactions in real time, without labeling any of the interactants (e.g., BIACore). Changes in the mass at the binding surface (indicative of a binding event) result in alterations of the refractive index of light near the surface (the optical phenomenon of surface plasmon resonance (SPR)), resulting in a detectable signal which can be used as an indication of real-time reactions between biological molecules.

[224] In one embodiment, the target gene product or the test substance is anchored onto a solid phase. The target gene product/test compound complexes anchored on the solid phase can be detected at the end of the reaction. Preferably, the target gene product can be anchored onto a solid surface, and the test compound, (which is not anchored), can be labeled, either directly or indirectly, with detectable labels discussed herein.

[225] It may be desirable to immobilize either 21953, an anti-21953 antibody or its target molecule to facilitate separation of complexed from uncomplexed forms of one or both of the proteins, as well as to accommodate automation of the assay. Binding of a test compound to a 21953 protein, or interaction of a 21953 protein with a target molecule in the presence and absence of a candidate compound, can be accomplished in any vessel suitable for containing the reactants. Examples of such vessels include microtiter plates, test tubes, and micro-centrifuge tubes. In one embodiment, a fusion protein can be provided which adds a domain that allows one or both of the proteins to be bound to a matrix. For example, glutathione-S-transferase/21953 fusion proteins or glutathione-S-transferase/target fusion proteins can be adsorbed onto glutathione sepharose beads (Sigma Chemical, St. Louis, MO) or glutathione derivatized microtiter plates, which are then combined with the test compound or the test compound and either the non-adsorbed target protein or 21953 protein, and the mixture incubated under conditions conducive to complex formation (e.g., at physiological conditions for salt and pH). Following incubation, the beads or microtiter plate wells are washed to remove any unbound components, the matrix immobilized in the case of beads, complex determined either directly or indirectly, for example, as described

above. Alternatively, the complexes can be dissociated from the matrix, and the level of 21953 binding or activity determined using standard techniques.

[226] Other techniques for immobilizing either a 21953 protein or a target molecule on matrices include using conjugation of biotin and streptavidin. Biotinylated 21953 protein or target molecules can be prepared from biotin-NHS (N-hydroxy-succinimide) using techniques known in the art (e.g., biotinylation kit, Pierce Chemicals, Rockford, IL), and immobilized in the wells of streptavidin-coated 96 well plates (Pierce Chemical).

[227] In order to conduct the assay, the non-immobilized component is added to the coated surface containing the anchored component. After the reaction is complete, unreacted components are removed (e.g., by washing) under conditions such that any complexes formed will remain immobilized on the solid surface. The detection of complexes anchored on the solid surface can be accomplished in a number of ways. Where the previously non-immobilized component is pre-labeled, the detection of label immobilized on the surface indicates that complexes were formed. Where the previously non-immobilized component is not pre-labeled, an indirect label can be used to detect complexes anchored on the surface; e.g., using a labeled antibody specific for the immobilized component (the antibody, in turn, can be directly labeled or indirectly labeled with, e.g., a labeled anti-Ig antibody).

[228] In one embodiment, this assay is performed utilizing antibodies reactive with 21953 protein or target molecules but which do not interfere with binding of the 21953 protein to its target molecule. Such antibodies can be derivatized to the wells of the plate, and unbound target or 21953 protein trapped in the wells by antibody conjugation. Methods for detecting such complexes, in addition to those described above for the GST-immobilized complexes, include immunodetection of complexes using antibodies reactive with the 21953 protein or target molecule, as well as enzyme-linked assays which rely on detecting an enzymatic activity associated with the 21953 protein or target molecule.

[229] Alternatively, cell free assays can be conducted in a liquid phase. In such an assay, the reaction products are separated from unreacted components, by any of a number of standard techniques, including but not limited to: differential centrifugation (see, for example, Rivas, G., and Minton, A.P., (1993) *Trends Biochem Sci* 18:284-7); chromatography (gel filtration chromatography, ion-exchange chromatography); electrophoresis (see, e.g., Ausubel, F. *et al.*, eds. *Current Protocols in Molecular Biology* 1999, J. Wiley: New York.); and immunoprecipitation (see, for example, Ausubel, F. *et al.*,

eds. (1999) *Current Protocols in Molecular Biology*, J. Wiley: New York). Such resins and chromatographic techniques are known to one skilled in the art (see, e.g., Heegaard, N.H., (1998) *J Mol Recognit* 11:141-8; Hage, D.S., and Tweed, S.A. (1997) *J Chromatogr B Biomed Sci Appl.* 699:499-525). Further, fluorescence energy transfer may also be conveniently utilized, as described herein, to detect binding without further purification of the complex from solution.

[230] In a preferred embodiment, the assay includes contacting the 21953 protein or biologically active portion thereof with a known compound which binds 21953 to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with a 21953 protein, wherein determining the ability of the test compound to interact with a 21953 protein includes determining the ability of the test compound to preferentially bind to 21953 or biologically active portion thereof, or to modulate the activity of a target molecule, as compared to the known compound.

[231] The target gene products of the invention can, *in vivo*, interact with one or more cellular or extracellular macromolecules, such as proteins. For the purposes of this discussion, such cellular and extracellular macromolecules are referred to herein as "binding partners." Compounds that disrupt such interactions can be useful in regulating the activity of the target gene product. Such compounds can include, but are not limited to molecules such as antibodies, peptides, and small molecules. The preferred target genes/products for use in this embodiment are the 21953 genes herein identified. In an alternative embodiment, the invention provides methods for determining the ability of the test compound to modulate the activity of a 21953 protein through modulation of the activity of a downstream effector of a 21953 target molecule. For example, the activity of the effector molecule on an appropriate target can be determined, or the binding of the effector to an appropriate target can be determined, as previously described.

[232] To identify compounds that interfere with the interaction between the target gene product and its cellular or extracellular binding partner(s), a reaction mixture containing the target gene product and the binding partner is prepared, under conditions and for a time sufficient, to allow the two products to form complex. In order to test an inhibitory agent, the reaction mixture is provided in the presence and absence of the test compound. The test compound can be initially included in the reaction mixture, or can be added at a time subsequent to the addition of the target gene and its cellular or extracellular binding partner. Control reaction mixtures are incubated without the test compound or with

a placebo. The formation of any complexes between the target gene product and the cellular or extracellular binding partner is then detected. The formation of a complex in the control reaction, but not in the reaction mixture containing the test compound, indicates that the compound interferes with the interaction of the target gene product and the interactive binding partner. Additionally, complex formation within reaction mixtures containing the test compound and normal target gene product can also be compared to complex formation within reaction mixtures containing the test compound and mutant target gene product. This comparison can be important in those cases wherein it is desirable to identify compounds that disrupt interactions of mutant but not normal target gene products.

[233] These assays can be conducted in a heterogeneous or homogeneous format. Heterogeneous assays involve anchoring either the target gene product or the binding partner onto a solid phase, and detecting complexes anchored on the solid phase at the end of the reaction. In homogeneous assays, the entire reaction is carried out in a liquid phase. In either approach, the order of addition of reactants can be varied to obtain different information about the compounds being tested. For example, test compounds that interfere with the interaction between the target gene products and the binding partners, e.g., by competition, can be identified by conducting the reaction in the presence of the test substance. Alternatively, test compounds that disrupt preformed complexes, e.g., compounds with higher binding constants that displace one of the components from the complex, can be tested by adding the test compound to the reaction mixture after complexes have been formed. The various formats are briefly described below.

[234] In a heterogeneous assay system, either the target gene product or the interactive cellular or extracellular binding partner, is anchored onto a solid surface (e.g., a microtiter plate), while the non-anchored species is labeled, either directly or indirectly. The anchored species can be immobilized by non-covalent or covalent attachments. Alternatively, an immobilized antibody specific for the species to be anchored can be used to anchor the species to the solid surface.

[235] In order to conduct the assay, the partner of the immobilized species is exposed to the coated surface with or without the test compound. After the reaction is complete, unreacted components are removed (e.g., by washing) and any complexes formed will remain immobilized on the solid surface. Where the non-immobilized species is pre-labeled, the detection of label immobilized on the surface indicates that complexes were formed. Where the non-immobilized species is not pre-labeled, an indirect label can be used

to detect complexes anchored on the surface; e.g., using a labeled antibody specific for the initially non-immobilized species (the antibody, in turn, can be directly labeled or indirectly labeled with, e.g., a labeled anti-Ig antibody). Depending upon the order of addition of reaction components, test compounds that inhibit complex formation or that disrupt preformed complexes can be detected.

[236] Alternatively, the reaction can be conducted in a liquid phase in the presence or absence of the test compound, the reaction products separated from unreacted components, and complexes detected; e.g., using an immobilized antibody specific for one of the binding components to anchor any complexes formed in solution, and a labeled antibody specific for the other partner to detect anchored complexes. Again, depending upon the order of addition of reactants to the liquid phase, test compounds that inhibit complex or that disrupt preformed complexes can be identified.

[237] In an alternate embodiment of the invention, a homogeneous assay can be used. For example, a preformed complex of the target gene product and the interactive cellular or extracellular binding partner product is prepared in that either the target gene products or their binding partners are labeled, but the signal generated by the label is quenched due to complex formation (see, e.g., U.S. Patent No. 4,109,496 that utilizes this approach for immunoassays). The addition of a test substance that competes with and displaces one of the species from the preformed complex will result in the generation of a signal above background. In this way, test substances that disrupt target gene product-binding partner interaction can be identified.

[238] In yet another aspect, the 21953 proteins can be used as "bait proteins" in a two-hybrid assay or three-hybrid assay (see, e.g., U.S. Patent No. 5,283,317; Zervos *et al.* (1993) *Cell* 72:223-232; Madura *et al.* (1993) *J. Biol. Chem.* 268:12046-12054; Bartel *et al.* (1993) *Biotechniques* 14:920-924; Iwabuchi *et al.* (1993) *Oncogene* 8:1693-1696; and Brent WO94/10300), to identify other proteins, which bind to or interact with 21953 ("21953-binding proteins" or "21953-bp") and are involved in 21953 activity. Such 21953-bps can be activators or inhibitors of signals by the 21953 proteins or 21953 targets as, for example, downstream elements of a 21953-mediated signaling pathway.

[239] The two-hybrid system is based on the modular nature of most transcription factors, which consist of separable DNA-binding and activation domains. Briefly, the assay utilizes two different DNA constructs. In one construct, the gene that codes for a 21953 protein is fused to a gene encoding the DNA binding domain of a known transcription factor

(e.g., GAL-4). In the other construct, a DNA sequence, from a library of DNA sequences, that encodes an unidentified protein ("prey" or "sample") is fused to a gene that codes for the activation domain of the known transcription factor. (Alternatively the 21953 protein can be fused to the activator domain.) If the "bait" and the "prey" proteins are able to interact, *in vivo*, forming a 21953-dependent complex, the DNA-binding and activation domains of the transcription factor are brought into close proximity. This proximity allows transcription of a reporter gene (e.g., lacZ) which is operably linked to a transcriptional regulatory site responsive to the transcription factor. Expression of the reporter gene can be detected and cell colonies containing the functional transcription factor can be isolated and used to obtain the cloned gene which encodes the protein which interacts with the 21953 protein.

[240] In another embodiment, modulators of 21953 expression are identified. For example, a cell or cell free mixture is contacted with a candidate compound and the expression of 21953 mRNA or protein evaluated relative to the level of expression of 21953 mRNA or protein in the absence of the candidate compound. When expression of 21953 mRNA or protein is greater in the presence of the candidate compound than in its absence, the candidate compound is identified as a stimulator of 21953 mRNA or protein expression. Alternatively, when expression of 21953 mRNA or protein is less (statistically significantly less) in the presence of the candidate compound than in its absence, the candidate compound is identified as an inhibitor of 21953 mRNA or protein expression. The level of 21953 mRNA or protein expression can be determined by methods described herein for detecting 21953 mRNA or protein.

[241] In another aspect, the invention pertains to a combination of two or more of the assays described herein. For example, a modulating agent can be identified using a cell-based or a cell free assay, and the ability of the agent to modulate the activity of a 21953 protein can be confirmed *in vivo*, e.g., in an animal such as an animal model for a cell proliferative or cell differentiative disorder, e.g., a cancer, e.g., a cancer of the lung, prostate, breast, or colon; an animal model for an immunological disorder; or an animal model for a neurological disorder.

[242] This invention further pertains to novel agents identified by the above-described screening assays. Accordingly, it is within the scope of this invention to further use an agent identified as described herein (e.g., a 21953 modulating agent, an antisense 21953 nucleic acid molecule, a 21953-specific antibody, or a 21953-binding partner) in an appropriate

animal model to determine the efficacy, toxicity, side effects, or mechanism of action, of treatment with such an agent. Furthermore, novel agents identified by the above-described screening assays can be used for treatments as described herein.

#### Detection Assays

[243] Portions or fragments of the nucleic acid sequences identified herein can be used as polynucleotide reagents. For example, these sequences can be used to: (i) map their respective genes on a chromosome e.g., to locate gene regions associated with genetic disease or to associate 21953 with a disease; (ii) identify an individual from a minute biological sample (tissue typing); and (iii) aid in forensic identification of a biological sample. These applications are described in the subsections below.

#### Chromosome Mapping

[244] The 21953 nucleotide sequences or portions thereof can be used to map the location of the 21953 genes on a chromosome. This process is called chromosome mapping. Chromosome mapping is useful in correlating the 21953 sequences with genes associated with disease.

[245] Briefly, 21953 genes can be mapped to chromosomes by preparing PCR primers (preferably 15-25 bp in length) from the 21953 nucleotide sequences. These primers can then be used for PCR screening of somatic cell hybrids containing individual human chromosomes. Only those hybrids containing the human gene corresponding to the 21953 sequences will yield an amplified fragment.

[246] A panel of somatic cell hybrids in which each cell line contains either a single human chromosome or a small number of human chromosomes, and a full set of mouse chromosomes, can allow easy mapping of individual genes to specific human chromosomes. (D'Eustachio P. *et al.* (1983) *Science* 220:919-924).

[247] Other mapping strategies e.g., *in situ* hybridization (described in Fan, Y. *et al.* (1990) *Proc. Natl. Acad. Sci. USA*, 87:6223-27), pre-screening with labeled flow-sorted chromosomes, and pre-selection by hybridization to chromosome specific cDNA libraries can be used to map 21953 to a chromosomal location.

[248] Fluorescence *in situ* hybridization (FISH) of a DNA sequence to a metaphase chromosomal spread can further be used to provide a precise chromosomal location in one step. The FISH technique can be used with a DNA sequence as short as 500 or 600 bases.

However, clones larger than 1,000 bases have a higher likelihood of binding to a unique chromosomal location with sufficient signal intensity for simple detection. Preferably 1,000 bases, and more preferably 2,000 bases will suffice to get good results at a reasonable amount of time. For a review of this technique, see Verma *et al.*, Human Chromosomes: A Manual of Basic Techniques ((1988) Pergamon Press, New York).

[249] Reagents for chromosome mapping can be used individually to mark a single chromosome or a single site on that chromosome, or panels of reagents can be used for marking multiple sites and/or multiple chromosomes. Reagents corresponding to noncoding regions of the genes actually are preferred for mapping purposes. Coding sequences are more likely to be conserved within gene families, thus increasing the chance of cross hybridizations during chromosomal mapping.

[250] Once a sequence has been mapped to a precise chromosomal location, the physical position of the sequence on the chromosome can be correlated with genetic map data. (Such data are found, for example, in V. McKusick, Mendelian Inheritance in Man, available on-line through Johns Hopkins University Welch Medical Library). The relationship between a gene and a disease, mapped to the same chromosomal region, can then be identified through linkage analysis (co-inheritance of physically adjacent genes), described in, for example, Egeland, J. *et al.* (1987) *Nature*, 325:783-787.

[251] Moreover, differences in the DNA sequences between individuals affected and unaffected with a disease associated with the 21953 gene, can be determined. If a mutation is observed in some or all of the affected individuals but not in any unaffected individuals, then the mutation is likely to be the causative agent of the particular disease. Comparison of affected and unaffected individuals generally involves first looking for structural alterations in the chromosomes, such as deletions or translocations that are visible from chromosome spreads or detectable using PCR based on that DNA sequence. Ultimately, complete sequencing of genes from several individuals can be performed to confirm the presence of a mutation and to distinguish mutations from polymorphisms.

#### Tissue Typing

[252] 21953 sequences can be used to identify individuals from biological samples using, e.g., restriction fragment length polymorphism (RFLP). In this technique, an individual's genomic DNA is digested with one or more restriction enzymes, the fragments separated, e.g., in a Southern blot, and probed to yield bands for identification. The

sequences of the present invention are useful as additional DNA markers for RFLP (described in U.S. Patent 5,272,057).

[253] Furthermore, the sequences of the present invention can also be used to determine the actual base-by-base DNA sequence of selected portions of an individual's genome. Thus, the 21953 nucleotide sequences described herein can be used to prepare two PCR primers from the 5' and 3' ends of the sequences. These primers can then be used to amplify an individual's DNA and subsequently sequence it. Panels of corresponding DNA sequences from individuals, prepared in this manner, can provide unique individual identifications, as each individual will have a unique set of such DNA sequences due to allelic differences.

[254] Allelic variation occurs to some degree in the coding regions of these sequences, and to a greater degree in the noncoding regions. Each of the sequences described herein can, to some degree, be used as a standard against which DNA from an individual can be compared for identification purposes. Because greater numbers of polymorphisms occur in the noncoding regions, fewer sequences are necessary to differentiate individuals. The noncoding sequences of SEQ ID NO:1 can provide positive individual identification with a panel of perhaps 10 to 1,000 primers which each yield a noncoding amplified sequence of 100 bases. If predicted coding sequences, such as those in SEQ ID NO:3 are used, a more appropriate number of primers for positive individual identification would be 500-2,000.

[255] If a panel of reagents from 21953 nucleotide sequences described herein is used to generate a unique identification database for an individual, those same reagents can later be used to identify tissue from that individual. Using the unique identification database, positive identification of the individual, living or dead, can be made from extremely small tissue samples.

#### Use of Partial 21953 Sequences in Forensic Biology

[256] DNA-based identification techniques can also be used in forensic biology. To make such an identification, PCR technology can be used to amplify DNA sequences taken from very small biological samples such as tissues, e.g., hair or skin, or body fluids, e.g., blood, saliva, or semen found at a crime scene. The amplified sequence can then be compared to a standard, thereby allowing identification of the origin of the biological sample.

[257] The sequences of the present invention can be used to provide polynucleotide reagents, e.g., PCR primers, targeted to specific loci in the human genome, which can enhance the reliability of DNA-based forensic identifications by, for example, providing another "identification marker" (i.e. another DNA sequence that is unique to a particular individual). As mentioned above, actual base sequence information can be used for identification as an accurate alternative to patterns formed by restriction enzyme generated fragments. Sequences targeted to noncoding regions of SEQ ID NO:1 (e.g., fragments derived from the noncoding regions of SEQ ID NO:1 having a length of at least 20 bases, preferably at least 30 bases) are particularly appropriate for this use.

[258] The 21953 nucleotide sequences described herein can further be used to provide polynucleotide reagents, e.g., labeled or labelable probes which can be used in, for example, an *in situ* hybridization technique, to identify a specific tissue. This can be very useful in cases where a forensic pathologist is presented with a tissue of unknown origin. Panels of such 21953 probes can be used to identify tissue by species and/or by organ type.

[259] In a similar fashion, these reagents, e.g., 21953 primers or probes can be used to screen tissue culture for contamination (i.e. screen for the presence of a mixture of different types of cells in a culture).

#### Predictive Medicine

[260] The present invention also pertains to the field of predictive medicine in which diagnostic assays, prognostic assays, and monitoring clinical trials are used for prognostic (predictive) purposes to thereby treat an individual.

[261] Generally, the invention provides, a method of determining if a subject is at risk for a disorder related to a lesion in or the misexpression of a gene which encodes 21953.

[262] Such disorders include, e.g., a disorder associated with the misexpression of 21953 gene; a disorder of cell proliferation (such as lung, breast, colon, prostate, or ovarian cancer) or of the nervous system.

[263] The method includes one or more of the following:

[264] detecting, in a tissue of the subject, the presence or absence of a mutation which affects the expression of the 21953 gene, or detecting the presence or absence of a mutation in a region which controls the expression of the gene, e.g., a mutation in the 5' control region;

[265] detecting, in a tissue of the subject, the presence or absence of a mutation which alters the structure of the 21953 gene;

[266] detecting, in a tissue of the subject, the misexpression of the 21953 gene, at the mRNA level, e.g., detecting a non-wild type level of a mRNA ;

[267] detecting, in a tissue of the subject, the misexpression of the gene, at the protein level, e.g., detecting a non-wild type level of a 21953 polypeptide.

[268] In preferred embodiments the method includes: ascertaining the existence of at least one of: a deletion of one or more nucleotides from the 21953 gene; an insertion of one or more nucleotides into the gene, a point mutation, e.g., a substitution of one or more nucleotides of the gene, a gross chromosomal rearrangement of the gene, e.g., a translocation, inversion, or deletion.

[269] For example, detecting the genetic lesion can include: (i) providing a probe/primer including an oligonucleotide containing a region of nucleotide sequence which hybridizes to a sense or antisense sequence from SEQ ID NO:1, or naturally occurring mutants thereof or 5' or 3' flanking sequences naturally associated with the 21953 gene; (ii) exposing the probe/primer to nucleic acid of the tissue; and detecting, by hybridization, e.g., *in-situ* hybridization, of the probe/primer to the nucleic acid, the presence or absence of the genetic lesion.

[270] In preferred embodiments detecting the misexpression includes ascertaining the existence of at least one of: an alteration in the level of a messenger RNA transcript of the 21953 gene; the presence of a non-wild type splicing pattern of a messenger RNA transcript of the gene; or a non-wild type level of 21953.

[271] Methods of the invention can be used prenatally or to determine if a subject's offspring will be at risk for a disorder.

[272] In preferred embodiments the method includes determining the structure of a 21953 gene, an abnormal structure being indicative of risk for the disorder.

[273] In preferred embodiments the method includes contacting a sample from the subject with an antibody to the 21953 protein or a nucleic acid, which hybridizes specifically with the gene. These and other embodiments are discussed below.

**Diagnostic and Prognostic Assays**

[274] Diagnostic and prognostic assays of the invention include method for assessing the expression level of 21953 molecules and for identifying variations and mutations in the sequence of 21953 molecules.

[275] *Expression Monitoring and Profiling.* The presence, level, or absence of 21953 protein or nucleic acid in a biological sample can be evaluated by obtaining a biological sample from a test subject and contacting the biological sample with a compound or an agent capable of detecting 21953 protein or nucleic acid (e.g., mRNA, genomic DNA) that encodes 21953 protein such that the presence of 21953 protein or nucleic acid is detected in the biological sample. The term "biological sample" includes tissues, cells and biological fluids isolated from a subject, as well as tissues, cells and fluids present within a subject. A preferred biological sample is serum. The level of expression of the 21953 gene can be measured in a number of ways, including, but not limited to: measuring the mRNA encoded by the 21953 genes; measuring the amount of protein encoded by the 21953 genes; or measuring the activity of the protein encoded by the 21953 genes.

[276] The level of mRNA corresponding to the 21953 gene in a cell can be determined both by *in situ* and by *in vitro* formats.

[277] The isolated mRNA can be used in hybridization or amplification assays that include, but are not limited to, Southern or Northern analyses, polymerase chain reaction analyses and probe arrays. One preferred diagnostic method for the detection of mRNA levels involves contacting the isolated mRNA with a nucleic acid molecule (probe) that can hybridize to the mRNA encoded by the gene being detected. The nucleic acid probe can be, for example, a full-length 21953 nucleic acid, such as the nucleic acid of SEQ ID NO:1, or a portion thereof, such as an oligonucleotide of at least 7, 15, 30, 50, 100, 250 or 500 nucleotides in length and sufficient to specifically hybridize under stringent conditions to 21953 mRNA or genomic DNA. The probe can be disposed on an address of an array, e.g., an array described below. Other suitable probes for use in the diagnostic assays are described herein.

[278] In one format, mRNA (or cDNA) is immobilized on a surface and contacted with the probes, for example by running the isolated mRNA on an agarose gel and transferring the mRNA from the gel to a membrane, such as nitrocellulose. In an alternative format, the probes are immobilized on a surface and the mRNA (or cDNA) is contacted with

the probes, for example, in a two-dimensional gene chip array described below. A skilled artisan can adapt known mRNA detection methods for use in detecting the level of mRNA encoded by the 21953 genes.

[279] The level of mRNA in a sample that is encoded by one of 21953 can be evaluated with nucleic acid amplification, e.g., by rtPCR (Mullis (1987) U.S. Patent No. 4,683,202), ligase chain reaction (Barany (1991) *Proc. Natl. Acad. Sci. USA* 88:189-193), self sustained sequence replication (Guatelli *et al.*, (1990) *Proc. Natl. Acad. Sci. USA* 87:1874-1878), transcriptional amplification system (Kwoh *et al.*, (1989), *Proc. Natl. Acad. Sci. USA* 86:1173-1177), Q-Beta Replicase (Lizardi *et al.*, (1988) *Bio/Technology* 6:1197), rolling circle replication (Lizardi *et al.*, U.S. Patent No. 5,854,033) or any other nucleic acid amplification method, followed by the detection of the amplified molecules using techniques known in the art. As used herein, amplification primers are defined as being a pair of nucleic acid molecules that can anneal to 5' or 3' regions of a gene (plus and minus strands, respectively, or vice-versa) and contain a short region in between. In general, amplification primers are from about 10 to 30 nucleotides in length and flank a region from about 50 to 200 nucleotides in length. Under appropriate conditions and with appropriate reagents, such primers permit the amplification of a nucleic acid molecule comprising the nucleotide sequence flanked by the primers.

[280] For *in situ* methods, a cell or tissue sample can be prepared/processed and immobilized on a support, typically a glass slide, and then contacted with a probe that can hybridize to mRNA that encodes the 21953 gene being analyzed.

[281] In another embodiment, the methods further contacting a control sample with a compound or agent capable of detecting 21953 mRNA, or genomic DNA, and comparing the presence of 21953 mRNA or genomic DNA in the control sample with the presence of 21953 mRNA or genomic DNA in the test sample. In still another embodiment, serial analysis of gene expression, as described in U.S. Patent No. 5,695,937, is used to detect 21953 transcript levels.

[282] A variety of methods can be used to determine the level of protein encoded by 21953. In general, these methods include contacting an agent that selectively binds to the protein, such as an antibody with a sample, to evaluate the level of protein in the sample. In a preferred embodiment, the antibody bears a detectable label. Antibodies can be polyclonal, or more preferably, monoclonal. An intact antibody, or a fragment thereof (e.g., Fab or F(ab')<sub>2</sub>) can be used. The term "labeled", with regard to the probe or antibody, is

intended to encompass direct labeling of the probe or antibody by coupling (i.e., physically linking) a detectable substance to the probe or antibody, as well as indirect labeling of the probe or antibody by reactivity with a detectable substance. Examples of detectable substances are provided herein.

[283] The detection methods can be used to detect 21953 protein in a biological sample *in vitro* as well as *in vivo*. *In vitro* techniques for detection of 21953 protein include enzyme linked immunosorbent assays (ELISAs), immunoprecipitations, immunofluorescence, enzyme immunoassay (EIA), radioimmunoassay (RIA), and Western blot analysis. *In vivo* techniques for detection of 21953 protein include introducing into a subject a labeled anti-21953 antibody. For example, the antibody can be labeled with a radioactive marker whose presence and location in a subject can be detected by standard imaging techniques. In another embodiment, the sample is labeled, e.g., biotinylated and then contacted to the antibody, e.g., an anti-21953 antibody positioned on an antibody array (as described below). The sample can be detected, e.g., with avidin coupled to a fluorescent label.

[284] In another embodiment, the methods further include contacting the control sample with a compound or agent capable of detecting 21953 protein, and comparing the presence of 21953 protein in the control sample with the presence of 21953 protein in the test sample.

[285] The invention also includes kits for detecting the presence of 21953 in a biological sample. For example, the kit can include a compound or agent capable of detecting 21953 protein or mRNA in a biological sample; and a standard. The compound or agent can be packaged in a suitable container. The kit can further comprise instructions for using the kit to detect 21953 protein or nucleic acid.

[286] For antibody-based kits, the kit can include: (1) a first antibody (e.g., attached to a solid support) which binds to a polypeptide corresponding to a marker of the invention; and, optionally, (2) a second, different antibody which binds to either the polypeptide or the first antibody and is conjugated to a detectable agent.

[287] For oligonucleotide-based kits, the kit can include: (1) an oligonucleotide, e.g., a detectably labeled oligonucleotide, which hybridizes to a nucleic acid sequence encoding a polypeptide corresponding to a marker of the invention or (2) a pair of primers useful for amplifying a nucleic acid molecule corresponding to a marker of the invention. The kit can also include a buffering agent, a preservative, or a protein stabilizing agent. The kit can

also includes components necessary for detecting the detectable agent (e.g., an enzyme or a substrate). The kit can also contain a control sample or a series of control samples which can be assayed and compared to the test sample contained. Each component of the kit can be enclosed within an individual container and all of the various containers can be within a single package, along with instructions for interpreting the results of the assays performed using the kit.

[288] The diagnostic methods described herein can identify subjects having, or at risk of developing, a disease or disorder associated with misexpressed or aberrant or unwanted 21953 expression or activity. As used herein, the term "unwanted" includes an unwanted phenomenon involved in a biological response such as a cell proliferative or cell differentiative disorder, e.g., a cancer, e.g., a cancer of the lung, prostate, breast, or colon or deregulated cell proliferation.

[289] In one embodiment, a disease or disorder associated with aberrant or unwanted 21953 expression or activity is identified. A test sample is obtained from a subject and 21953 protein or nucleic acid (e.g., mRNA or genomic DNA) is evaluated, wherein the level, e.g., the presence or absence, of 21953 protein or nucleic acid is diagnostic for a subject having or at risk of developing a disease or disorder associated with aberrant or unwanted 21953 expression or activity. As used herein, a "test sample" refers to a biological sample obtained from a subject of interest, including a biological fluid (e.g., serum), cell sample, or tissue.

[290] The prognostic assays described herein can be used to determine whether a subject can be administered an agent (e.g., an agonist, antagonist, peptidomimetic, protein, peptide, nucleic acid, small molecule, or other drug candidate) to treat a disease or disorder associated with aberrant or unwanted 21953 expression or activity. For example, such methods can be used to determine whether a subject can be effectively treated with an agent for a cell a cell proliferative or cell differentiative disorder, e.g., a cancer, e.g., a cancer of the lung, prostate, breast, or colon disorder.

[291] In another aspect, the invention features a computer medium having a plurality of digitally encoded data records. Each data record includes a value representing the level of expression of 21953 in a sample, and a descriptor of the sample. The descriptor of the sample can be an identifier of the sample, a subject from which the sample was derived (e.g., a patient), a diagnosis, or a treatment (e.g., a preferred treatment). In a preferred embodiment, the data record further includes values representing the level of expression of

genes other than 21953 (e.g., other genes associated with a 21953-disorder, or other genes on an array). The data record can be structured as a table, e.g., a table that is part of a database such as a relational database (e.g., a SQL database of the Oracle or Sybase database environments).

[292] Also featured is a method of evaluating a sample. The method includes providing a sample, e.g., from the subject, and determining a gene expression profile of the sample, wherein the profile includes a value representing the level of 21953 expression. The method can further include comparing the value or the profile (i.e., multiple values) to a reference value or reference profile. The gene expression profile of the sample can be obtained by any of the methods described herein (e.g., by providing a nucleic acid from the sample and contacting the nucleic acid to an array). The method can be used to diagnose a cell proliferative or cell differentiative disorder, e.g., a cancer, e.g., a cancer of the lung, in a subject wherein an increase in 21953 expression is an indication that the subject has or is disposed to having a cell proliferative or cell differentiative disorder, e.g., a cancer, e.g., a cancer of the lung. The method can be used to monitor a treatment for a cell proliferative or cell differentiative disorder, e.g., a cancer, e.g., a cancer of the lung, prostate, breast, or colon in a subject. For example, the gene expression profile can be determined for a sample from a subject undergoing treatment. The profile can be compared to a reference profile or to a profile obtained from the subject prior to treatment or prior to onset of the disorder (see, e.g., Golub *et al.* (1999) *Science* 286:531).

[293] In yet another aspect, the invention features a method of evaluating a test compound (see also, "Screening Assays", above). The method includes providing a cell and a test compound; contacting the test compound to the cell; obtaining a subject expression profile for the contacted cell; and comparing the subject expression profile to one or more reference profiles. The profiles include a value representing the level of 21953 expression. In a preferred embodiment, the subject expression profile is compared to a target profile, e.g., a profile for a normal cell or for desired condition of a cell. The test compound is evaluated favorably if the subject expression profile is more similar to the target profile than an expression profile obtained from an uncontacted cell.

[294] In another aspect, the invention features, a method of evaluating a subject. The method includes: a) obtaining a sample from a subject, e.g., from a caregiver, e.g., a caregiver who obtains the sample from the subject; b) determining a subject expression profile for the sample. Optionally, the method further includes either or both of steps: c)

comparing the subject expression profile to one or more reference expression profiles; and d) selecting the reference profile most similar to the subject reference profile. The subject expression profile and the reference profiles include a value representing the level of 21953 expression. A variety of routine statistical measures can be used to compare two reference profiles. One possible metric is the length of the distance vector that is the difference between the two profiles. Each of the subject and reference profile is represented as a multi-dimensional vector, wherein each dimension is a value in the profile.

[295] The method can further include transmitting a result to a caregiver. The result can be the subject expression profile, a result of a comparison of the subject expression profile with another profile, a most similar reference profile, or a descriptor of any of the aforementioned. The result can be transmitted across a computer network, e.g., the result can be in the form of a computer transmission, e.g., a computer data signal embedded in a carrier wave.

[296] Also featured is a computer medium having executable code for effecting the following steps: receive a subject expression profile; access a database of reference expression profiles; and either i) select a matching reference profile most similar to the subject expression profile; or ii) determine at least one comparison score for the similarity of the subject expression profile to at least one reference profile. The subject expression profile, and the reference expression profiles each include a value representing the level of 21953 expression.

#### Arrays and Uses Thereof

[297] In another aspect, the invention features an array that includes a substrate having a plurality of addresses. At least one address of the plurality includes a capture probe that binds specifically to a 21953 molecule (e.g., a 21953 nucleic acid or a 21953 polypeptide). The array can have a density of at least than 10, 50, 100, 200, 500, 1,000, 2,000, or 10,000 or more addresses/cm<sup>2</sup>, and ranges between. In a preferred embodiment, the plurality of addresses includes at least 10, 100, 500, 1,000, 5,000, 10,000, 50,000 addresses. In a preferred embodiment, the plurality of addresses includes equal to or less than 10, 100, 500, 1,000, 5,000, 10,000, or 50,000 addresses. The substrate can be a two-dimensional substrate such as a glass slide, a wafer (e.g., silica or plastic), a mass spectroscopy plate, or a three-dimensional substrate such as a gel pad. Addresses in addition to address of the plurality can be disposed on the array.

[298] In a preferred embodiment, at least one address of the plurality includes a nucleic acid capture probe that hybridizes specifically to a 21953 nucleic acid, e.g., the sense or anti-sense strand. In one preferred embodiment, a subset of addresses of the plurality of addresses has a nucleic acid capture probe for 21953. Each address of the subset can include a capture probe that hybridizes to a different region of a 21953 nucleic acid. In another preferred embodiment, addresses of the subset include a capture probe for a 21953 nucleic acid. Each address of the subset is unique, overlapping, and complementary to a different variant of 21953 (e.g., an allelic variant, or all possible hypothetical variants). The array can be used to sequence 21953 by hybridization (see, e.g., U.S. Patent No. 5,695,940).

[299] An array can be generated by various methods, e.g., by photolithographic methods (see, e.g., U.S. Patent Nos. 5,143,854; 5,510,270; and 5,527,681), mechanical methods (e.g., directed-flow methods as described in U.S. Patent No. 5,384,261), pin-based methods (e.g., as described in U.S. Pat. No. 5,288,514), and bead-based techniques (e.g., as described in PCT US/93/04145).

[300] In another preferred embodiment, at least one address of the plurality includes a polypeptide capture probe that binds specifically to a 21953 polypeptide or fragment thereof. The polypeptide can be a naturally-occurring interaction partner of 21953 polypeptide. Preferably, the polypeptide is an antibody, e.g., an antibody described herein (see "Anti-21953 Antibodies," above), such as a monoclonal antibody or a single-chain antibody.

[301] In another aspect, the invention features a method of analyzing the expression of 21953. The method includes providing an array as described above; contacting the array with a sample and detecting binding of a 21953-molecule (e.g., nucleic acid or polypeptide) to the array. In a preferred embodiment, the array is a nucleic acid array. Optionally the method further includes amplifying nucleic acid from the sample prior or during contact with the array.

[302] In another embodiment, the array can be used to assay gene expression in a tissue to ascertain tissue specificity of genes in the array, particularly the expression of 21953. If a sufficient number of diverse samples is analyzed, clustering (e.g., hierarchical clustering, k-means clustering, Bayesian clustering and the like) can be used to identify other genes which are co-regulated with 21953. For example, the array can be used for the quantitation of the expression of multiple genes. Thus, not only tissue specificity, but also the level of expression of a battery of genes in the tissue is ascertained. Quantitative data

can be used to group (e.g., cluster) genes on the basis of their tissue expression *per se* and level of expression in that tissue.

[303] For example, array analysis of gene expression can be used to assess the effect of cell-cell interactions on 21953 expression. A first tissue can be perturbed and nucleic acid from a second tissue that interacts with the first tissue can be analyzed. In this context, the effect of one cell type on another cell type in response to a biological stimulus can be determined, e.g., to monitor the effect of cell-cell interaction at the level of gene expression.

[304] In another embodiment, cells are contacted with a therapeutic agent. The expression profile of the cells is determined using the array, and the expression profile is compared to the profile of like cells not contacted with the agent. For example, the assay can be used to determine or analyze the molecular basis of an undesirable effect of the therapeutic agent. If an agent is administered therapeutically to treat one cell type but has an undesirable effect on another cell type, the invention provides an assay to determine the molecular basis of the undesirable effect and thus provides the opportunity to co-administer a counteracting agent or otherwise treat the undesired effect. Similarly, even within a single cell type, undesirable biological effects can be determined at the molecular level. Thus, the effects of an agent on expression of other than the target gene can be ascertained and counteracted.

[305] In another embodiment, the array can be used to monitor expression of one or more genes in the array with respect to time. For example, samples obtained from different time points can be probed with the array. Such analysis can identify and/or characterize the development of a 21953-associated disease or disorder, and processes, such as a cellular transformation associated with a 21953-associated disease or disorder. The method can also evaluate the treatment and/or progression of a 21953-associated disease or disorder.

[306] The array is also useful for ascertaining differential expression patterns of one or more genes in normal and abnormal cells. This provides a battery of genes (e.g., including 21953) that could serve as a molecular target for diagnosis or therapeutic intervention.

[307] In another aspect, the invention features an array having a plurality of addresses. Each address of the plurality includes a unique polypeptide. At least one address of the plurality has disposed thereon a 21953 polypeptide or fragment thereof. Methods of producing polypeptide arrays are described in the art, e.g., in De Wildt *et al.* (2000). *Nature Biotech.* 18, 989-994; Lueking *et al.* (1999). *Anal. Biochem.* 270, 103-111; Ge, H. (2000). *Nucleic Acids Res.* 28, e3, I-VII; MacBeath, G., and Schreiber, S.L. (2000). *Science* 289,

1760-1763; and WO 99/51773A1. In a preferred embodiment, each addresses of the plurality has disposed thereon a polypeptide at least 60, 70, 80, 85, 90, 95 or 99 % identical to a 21953 polypeptide or fragment thereof. For example, multiple variants of a 21953 polypeptide (e.g., encoded by allelic variants, site-directed mutants, random mutants, or combinatorial mutants) can be disposed at individual addresses of the plurality. Addresses in addition to the address of the plurality can be disposed on the array.

[308] The polypeptide array can be used to detect a 21953 binding compound, e.g., an antibody in a sample from a subject with specificity for a 21953 polypeptide or the presence of a 21953-binding protein or ligand.

[309] The array is also useful for ascertaining the effect of the expression of a gene on the expression of other genes in the same cell or in different cells (e.g., ascertaining the effect of 21953 expression on the expression of other genes). This provides, for example, for a selection of alternate molecular targets for therapeutic intervention if the ultimate or downstream target cannot be regulated.

[310] In another aspect, the invention features a method of analyzing a plurality of probes. The method is useful, e.g., for analyzing gene expression. The method includes: providing a two dimensional array having a plurality of addresses, each address of the plurality being positionally distinguishable from each other address of the plurality having a unique capture probe, e.g., wherein the capture probes are from a cell or subject which express 21953 or from a cell or subject in which a 21953 mediated response has been elicited, e.g., by contact of the cell with 21953 nucleic acid or protein, or administration to the cell or subject 21953 nucleic acid or protein; providing a two dimensional array having a plurality of addresses, each address of the plurality being positionally distinguishable from each other address of the plurality, and each address of the plurality having a unique capture probe, e.g., wherein the capture probes are from a cell or subject which does not express 21953 (or does not express as highly as in the case of the 21953 positive plurality of capture probes) or from a cell or subject which in which a 21953 mediated response has not been elicited (or has been elicited to a lesser extent than in the first sample); contacting the array with one or more inquiry probes (which is preferably other than a 21953 nucleic acid, polypeptide, or antibody), and thereby evaluating the plurality of capture probes. Binding, e.g., in the case of a nucleic acid, hybridization with a capture probe at an address of the plurality, is detected, e.g., by signal generated from a label attached to the nucleic acid, polypeptide, or antibody.

[311] In another aspect, the invention features a method of analyzing a plurality of probes or a sample. The method is useful, e.g., for analyzing gene expression. The method includes: providing a two dimensional array having a plurality of addresses, each address of the plurality being positionally distinguishable from each other address of the plurality having a unique capture probe, contacting the array with a first sample from a cell or subject which express or mis-express 21953 or from a cell or subject in which a 21953-mediated response has been elicited, e.g., by contact of the cell with 21953 nucleic acid or protein, or administration to the cell or subject 21953 nucleic acid or protein; providing a two dimensional array having a plurality of addresses, each address of the plurality being positionally distinguishable from each other address of the plurality, and each address of the plurality having a unique capture probe, and contacting the array with a second sample from a cell or subject which does not express 21953 (or does not express as highly as in the case of the 21953 positive plurality of capture probes) or from a cell or subject which in which a 21953-mediated response has not been elicited (or has been elicited to a lesser extent than in the first sample); and comparing the binding of the first sample with the binding of the second sample. Binding, e.g., in the case of a nucleic acid, hybridization with a capture probe at an address of the plurality, is detected, e.g., by signal generated from a label attached to the nucleic acid, polypeptide, or antibody. The same array can be used for both samples or different arrays can be used. If different arrays are used the plurality of addresses with capture probes should be present on both arrays.

[312] In another aspect, the invention features a method of analyzing 21953, e.g., analyzing structure, function, or relatedness to other nucleic acid or amino acid sequences. The method includes: providing a 21953 nucleic acid or amino acid sequence; comparing the 21953 sequence with one or more preferably a plurality of sequences from a collection of sequences, e.g., a nucleic acid or protein sequence database; to thereby analyze 21953.

#### Detection of Variations or Mutations

[313] The methods of the invention can also be used to detect genetic alterations in a 21953 gene, thereby determining if a subject with the altered gene is at risk for a disorder characterized by misregulation in 21953 protein activity or nucleic acid expression, such as a cell proliferative or cell differentiative disorder, e.g., a cancer, e.g., a cancer of the lung, prostate, breast, or colon disorder. In preferred embodiments, the methods include detecting, in a sample from the subject, the presence or absence of a genetic alteration

characterized by at least one of an alteration affecting the integrity of a gene encoding a 21953-protein, or the mis-expression of the 21953 gene. For example, such genetic alterations can be detected by ascertaining the existence of at least one of 1) a deletion of one or more nucleotides from a 21953 gene; 2) an addition of one or more nucleotides to a 21953 gene; 3) a substitution of one or more nucleotides of a 21953 gene, 4) a chromosomal rearrangement of a 21953 gene; 5) an alteration in the level of a messenger RNA transcript of a 21953 gene, 6) aberrant modification of a 21953 gene, such as of the methylation pattern of the genomic DNA, 7) the presence of a non-wild type splicing pattern of a messenger RNA transcript of a 21953 gene, 8) a non-wild type level of a 21953-protein, 9) allelic loss of a 21953 gene, and 10) inappropriate post-translational modification of a 21953-protein.

[314] An alteration can be detected without a probe/primer in a polymerase chain reaction, such as anchor PCR or RACE PCR, or, alternatively, in a ligation chain reaction (LCR), the latter of which can be particularly useful for detecting point mutations in the 21953-gene. This method can include the steps of collecting a sample of cells from a subject, isolating nucleic acid (e.g., genomic, mRNA or both) from the sample, contacting the nucleic acid sample with one or more primers which specifically hybridize to a 21953 gene under conditions such that hybridization and amplification of the 21953-gene (if present) occurs, and detecting the presence or absence of an amplification product, or detecting the size of the amplification product and comparing the length to a control sample. It is anticipated that PCR and/or LCR may be desirable to use as a preliminary amplification step in conjunction with any of the techniques used for detecting mutations described herein. Alternatively, other amplification methods described herein or known in the art can be used.

[315] In another embodiment, mutations in a 21953 gene from a sample cell can be identified by detecting alterations in restriction enzyme cleavage patterns. For example,

---

e.g., chip based arrays. Such arrays include a plurality of addresses, each of which is positionally distinguishable from the other. A different probe is located at each address of the plurality. A probe can be complementary to a region of a 21953 nucleic acid or a putative variant (e.g., allelic variant) thereof. A probe can have one or more mismatches to a region of a 21953 nucleic acid (e.g., a destabilizing mismatch). The arrays can have a high density of addresses, e.g., can contain hundreds or thousands of oligonucleotides probes (Cronin, M.T. *et al.* (1996) *Human Mutation* 7: 244-255; Kozal, M.J. *et al.* (1996) *Nature Medicine* 2: 753-759). For example, genetic mutations in 21953 can be identified in two-dimensional arrays containing light-generated DNA probes as described in Cronin, M.T. *et al. supra*. Briefly, a first hybridization array of probes can be used to scan through long stretches of DNA in a sample and control to identify base changes between the sequences by making linear arrays of sequential overlapping probes. This step allows the identification of point mutations. This step is followed by a second hybridization array that allows the characterization of specific mutations by using smaller, specialized probe arrays complementary to all variants or mutations detected. Each mutation array is composed of parallel probe sets, one complementary to the wild-type gene and the other complementary to the mutant gene.

[317] In yet another embodiment, any of a variety of sequencing reactions known in the art can be used to directly sequence the 21953 gene and detect mutations by comparing the sequence of the sample 21953 with the corresponding wild-type (control) sequence. Automated sequencing procedures can be utilized when performing the diagnostic assays ((1995) *Biotechniques* 19:448), including sequencing by mass spectrometry.

[318] Other methods for detecting mutations in the 21953 gene include methods in which protection from cleavage agents is used to detect mismatched bases in RNA/RNA or RNA/DNA heteroduplexes (Myers *et al.* (1985) *Science* 230:1242; Cotton *et al.* (1988) *Proc. Natl Acad Sci USA* 85:4397; Saleeba *et al.* (1992) *Methods Enzymol.* 217:286-295).

[319] In still another embodiment, the mismatch cleavage reaction employs one or more proteins that recognize mismatched base pairs in double-stranded DNA (so called "DNA mismatch repair" enzymes) in defined systems for detecting and mapping point mutations in 21953 cDNAs obtained from samples of cells. For example, the mutY enzyme of *E. coli* cleaves A at G/A mismatches and the thymidine DNA glycosylase from HeLa cells cleaves T at G/T mismatches (Hsu *et al.* (1994) *Carcinogenesis* 15:1657-1662; U.S. Patent No. 5,459,039).

[320] In other embodiments, alterations in electrophoretic mobility will be used to identify mutations in 21953 genes. For example, single strand conformation polymorphism (SSCP) may be used to detect differences in electrophoretic mobility between mutant and wild type nucleic acids (Orita *et al.* (1989) *Proc Natl. Acad. Sci USA*: 86:2766, see also Cotton (1993) *Mutat. Res.* 285:125-144; and Hayashi (1992) *Genet. Anal. Tech. Appl.* 9:73-79). Single-stranded DNA fragments of sample and control 21953 nucleic acids will be denatured and allowed to renature. The secondary structure of single-stranded nucleic acids varies according to sequence, the resulting alteration in electrophoretic mobility enables the detection of even a single base change. The DNA fragments may be labeled or detected with labeled probes. The sensitivity of the assay may be enhanced by using RNA (rather than DNA), in which the secondary structure is more sensitive to a change in sequence. In a preferred embodiment, the subject method utilizes heteroduplex analysis to separate double stranded heteroduplex molecules on the basis of changes in electrophoretic mobility (Keen *et al.* (1991) *Trends Genet* 7:5).

[321] In yet another embodiment, the movement of mutant or wild-type fragments in polyacrylamide gels containing a gradient of denaturant is assayed using denaturing gradient gel electrophoresis (DGGE) (Myers *et al.* (1985) *Nature* 313:495). When DGGE is used as the method of analysis, DNA will be modified to insure that it does not completely denature, for example by adding a GC clamp of approximately 40 bp of high-melting GC-rich DNA by PCR. In a further embodiment, a temperature gradient is used in place of a denaturing gradient to identify differences in the mobility of control and sample DNA (Rosenbaum and Reissner (1987) *Biophys Chem* 265:12753).

[322] Examples of other techniques for detecting point mutations include, but are not limited to, selective oligonucleotide hybridization, selective amplification, or selective primer extension (Saiki *et al.* (1986) *Nature* 324:163); Saiki *et al.* (1989) *Proc. Natl. Acad. Sci USA* 86:6230). A further method of detecting point mutations is the chemical ligation of oligonucleotides as described in Xu *et al.* ((2001) *Nature Biotechnol.* 19:148). Adjacent oligonucleotides, one of which selectively anneals to the query site, are ligated together if the nucleotide at the query site of the sample nucleic acid is complementary to the query oligonucleotide; ligation can be monitored, e.g., by fluorescent dyes coupled to the oligonucleotides.

[323] Alternatively, allele specific amplification technology that depends on selective PCR amplification may be used in conjunction with the instant invention. Oligonucleotides

used as primers for specific amplification may carry the mutation of interest in the center of the molecule (so that amplification depends on differential hybridization) (Gibbs *et al.* (1989) *Nucleic Acids Res.* 17:2437-2448) or at the extreme 3' end of one primer where, under appropriate conditions, mismatch can prevent, or reduce polymerase extension (Prossner (1993) *Tibtech* 11:238). In addition it may be desirable to introduce a novel restriction site in the region of the mutation to create cleavage-based detection (Gasparini *et al.* (1992) *Mol. Cell Probes* 6:1). It is anticipated that in certain embodiments amplification may also be performed using Taq ligase for amplification (Barany (1991) *Proc. Natl. Acad. Sci USA* 88:189). In such cases, ligation will occur only if there is a perfect match at the 3' end of the 5' sequence making it possible to detect the presence of a known mutation at a specific site by looking for the presence or absence of amplification.

[324] In another aspect, the invention features a set of oligonucleotides. The set includes a plurality of oligonucleotides, each of which is at least partially complementary (e.g., at least 50%, 60%, 70%, 80%, 90%, 92%, 95%, 97%, 98%, or 99% complementary) to a 21953-nucleic acid.

[325] In a preferred embodiment the set includes a first and a second oligonucleotide. The first and second oligonucleotide can hybridize to the same or to different locations of SEQ ID NO:1 or the complement of SEQ ID NO:1. Different locations can be different but overlapping or non-overlapping on the same strand. The first and second oligonucleotide can hybridize to sites on the same or on different strands.

[326] The set can be useful, e.g., for identifying SNP's, or identifying specific alleles of 21953. In a preferred embodiment, each oligonucleotide of the set has a different nucleotide at an interrogation position. In one embodiment, the set includes two oligonucleotides, each complementary to a different allele at a locus, e.g., a biallelic or polymorphic locus.

[327] In another embodiment, the set includes four oligonucleotides, each having a different nucleotide (e.g., adenine, guanine, cytosine, or thymidine) at the interrogation position. The interrogation position can be a SNP or the site of a mutation. In another preferred embodiment, the oligonucleotides of the plurality are identical in sequence to one another (except for differences in length). The oligonucleotides can be provided with differential labels, such that an oligonucleotide that hybridizes to one allele provides a signal that is distinguishable from an oligonucleotide that hybridizes to a second allele. In still another embodiment, at least one of the oligonucleotides of the set has a nucleotide change

at a position in addition to a query position, e.g., a destabilizing mutation to decrease the  $T_m$  of the oligonucleotide. In another embodiment, at least one oligonucleotide of the set has a non-natural nucleotide, e.g., inosine. In a preferred embodiment, the oligonucleotides are attached to a solid support, e.g., to different addresses of an array or to different beads or nanoparticles.

[328] In a preferred embodiment the set of oligo nucleotides can be used to specifically amplify, e.g., by PCR, or detect, a 21953 nucleic acid.

[329] The methods described herein may be performed, for example, by utilizing pre-packaged diagnostic kits comprising at least one probe nucleic acid or antibody reagent described herein, which may be conveniently used, e.g., in clinical settings to diagnose patients exhibiting symptoms or family history of a disease or illness involving a 21953 gene.

#### Use of 21953 Molecules as Surrogate Markers

[330] The 21953 molecules of the invention are also useful as markers of disorders or disease states, as markers for precursors of disease states, as markers for predisposition of disease states, as markers of drug activity, or as markers of the pharmacogenomic profile of a subject. Using the methods described herein, the presence, absence and/or quantity of the 21953 molecules of the invention may be detected, and may be correlated with one or more biological states *in vivo*. For example, the 21953 molecules of the invention may serve as surrogate markers for one or more disorders or disease states or for conditions leading up to disease states. As used herein, a "surrogate marker" is an objective biochemical marker which correlates with the absence or presence of a disease or disorder, or with the progression of a disease or disorder (e.g., with the presence or absence of a tumor). The presence or quantity of such markers is independent of the disease. Therefore, these markers may serve to indicate whether a particular course of treatment is effective in lessening a disease state or disorder. Surrogate markers are of particular use when the presence or extent of a disease state or disorder is difficult to assess through standard methodologies (e.g., early stage tumors), or when an assessment of disease progression is desired before a potentially dangerous clinical endpoint is reached (e.g., an assessment of cardiovascular disease may be made using cholesterol levels as a surrogate marker, and an analysis of HIV infection may be made using HIV RNA levels as a surrogate marker, well in advance of the undesirable clinical outcomes of myocardial infarction or fully-developed

AIDS). Examples of the use of surrogate markers in the art include: Koomen *et al.* (2000) *J. Mass. Spectrom.* 35: 258-264; and James (1994) *AIDS Treatment News Archive* 209.

[331] The 21953 molecules of the invention are also useful as pharmacodynamic markers. As used herein, a "pharmacodynamic marker" is an objective biochemical marker which correlates specifically with drug effects. The presence or quantity of a pharmacodynamic marker is not related to the disease state or disorder for which the drug is being administered; therefore, the presence or quantity of the marker is indicative of the presence or activity of the drug in a subject. For example, a pharmacodynamic marker may be indicative of the concentration of the drug in a biological tissue, in that the marker is either expressed or transcribed or not expressed or transcribed in that tissue in relationship to the level of the drug. In this fashion, the distribution or uptake of the drug may be monitored by the pharmacodynamic marker. Similarly, the presence or quantity of the pharmacodynamic marker may be related to the presence or quantity of the metabolic product of a drug, such that the presence or quantity of the marker is indicative of the relative breakdown rate of the drug *in vivo*. Pharmacodynamic markers are of particular use in increasing the sensitivity of detection of drug effects, particularly when the drug is administered in low doses. Since even a small amount of a drug may be sufficient to activate multiple rounds of marker (e.g., a 21953 marker) transcription or expression, the amplified marker may be in a quantity which is more readily detectable than the drug itself. Also, the marker may be more easily detected due to the nature of the marker itself; for example, using the methods described herein, anti-21953 antibodies may be employed in an immune-based detection system for a 21953 protein marker, or 21953-specific radiolabeled probes may be used to detect a 21953 mRNA marker. Furthermore, the use of a pharmacodynamic marker may offer mechanism-based prediction of risk due to drug treatment beyond the range of possible direct observations. Examples of the use of pharmacodynamic markers in the art include: Matsuda *et al.* US 6,033,862; Hattis *et al.* (1991) *Env. Health Perspect.* 90: 229-238; Schentag (1999) *Am. J. Health-Syst. Pharm.* 56 Suppl. 3: S21-S24; and Nicolau (1999) *Am. J. Health-Syst. Pharm.* 56 Suppl. 3: S16-S20.

[332] The 21953 molecules of the invention are also useful as pharmacogenomic markers. As used herein, a "pharmacogenomic marker" is an objective biochemical marker which correlates with a specific clinical drug response or susceptibility in a subject (see, e.g., McLeod *et al.* (1999) *Eur. J. Cancer* 35:1650-1652). The presence or quantity of the pharmacogenomic marker is related to the predicted response of the subject to a specific

drug or class of drugs prior to administration of the drug. By assessing the presence or quantity of one or more pharmacogenomic markers in a subject, a drug therapy which is most appropriate for the subject, or which is predicted to have a greater degree of success, may be selected. For example, based on the presence or quantity of RNA, or protein (e.g., 21953 protein or RNA) for specific tumor markers in a subject, a drug or course of treatment may be selected that is optimized for the treatment of the specific tumor likely to be present in the subject. Similarly, the presence or absence of a specific sequence mutation in 21953 DNA may correlate 21953 drug response. The use of pharmacogenomic markers therefore permits the application of the most appropriate treatment for each subject without having to administer the therapy.

#### Pharmaceutical Compositions

[333] The nucleic acid and polypeptides, fragments thereof, as well as anti-21953 antibodies (also referred to herein as "active compounds") of the invention can be incorporated into pharmaceutical compositions. Such compositions typically include the nucleic acid molecule, protein, or antibody and a pharmaceutically acceptable carrier. As used herein the language "pharmaceutically acceptable carrier" includes solvents, dispersion media, coatings, antibacterial and antifungal agents, isotonic and absorption delaying agents, and the like, compatible with pharmaceutical administration. Supplementary active compounds can also be incorporated into the compositions.

[334] A pharmaceutical composition is formulated to be compatible with its intended route of administration. Examples of routes of administration include parenteral, e.g., intravenous, intradermal, subcutaneous, oral (e.g., inhalation), transdermal (topical), transmucosal, and rectal administration. Solutions or suspensions used for parenteral, intradermal, or subcutaneous application can include the following components: a sterile diluent such as water for injection, saline solution, fixed oils, polyethylene glycols, glycerine, propylene glycol or other synthetic solvents; antibacterial agents such as benzyl alcohol or methyl parabens; antioxidants such as ascorbic acid or sodium bisulfite; chelating agents such as ethylenediaminetetraacetic acid; buffers such as acetates, citrates or phosphates and agents for the adjustment of tonicity such as sodium chloride or dextrose. pH can be adjusted with acids or bases, such as hydrochloric acid or sodium hydroxide. The parenteral preparation can be enclosed in ampoules, disposable syringes or multiple dose vials made of glass or plastic.

[335] Pharmaceutical compositions suitable for injectable use include sterile aqueous solutions (where water soluble) or dispersions and sterile powders for the extemporaneous preparation of sterile injectable solutions or dispersion. For intravenous administration, suitable carriers include physiological saline, bacteriostatic water, Cremophor EL™ (BASF, Parsippany, NJ) or phosphate buffered saline (PBS). In all cases, the composition must be sterile and should be fluid to the extent that easy syringability exists. It should be stable under the conditions of manufacture and storage and must be preserved against the contaminating action of microorganisms such as bacteria and fungi. The carrier can be a solvent or dispersion medium containing, for example, water, ethanol, polyol (for example, glycerol, propylene glycol, and liquid polyethylene glycol, and the like), and suitable mixtures thereof. The proper fluidity can be maintained, for example, by the use of a coating such as lecithin, by the maintenance of the required particle size in the case of dispersion and by the use of surfactants. Prevention of the action of microorganisms can be achieved by various antibacterial and antifungal agents, for example, parabens, chlorobutanol, phenol, ascorbic acid, thimerosal, and the like. In many cases, it will be preferable to include isotonic agents, for example, sugars, polyalcohols such as manitol, sorbitol, sodium chloride in the composition. Prolonged absorption of the injectable compositions can be brought about by including in the composition an agent which delays absorption; for example, aluminum-monostearate and gelatin.

[336] Sterile injectable solutions can be prepared by incorporating the active compound in the required amount in an appropriate solvent with one or a combination of ingredients enumerated above, as required, followed by filtered sterilization. Generally, dispersions are prepared by incorporating the active compound into a sterile vehicle which contains a basic dispersion medium and the required other ingredients from those enumerated above. In the case of sterile powders for the preparation of sterile injectable solutions, the preferred methods of preparation are vacuum drying and freeze-drying which yields a powder of the active ingredient plus any additional desired ingredient from a previously sterile-filtered solution thereof.

[337] Oral compositions generally include an inert diluent or an edible carrier. For the purpose of oral therapeutic administration, the active compound can be incorporated with excipients and used in the form of tablets, troches, or capsules, e.g., gelatin capsules. Oral compositions can also be prepared using a fluid carrier for use as a mouthwash. Pharmaceutically compatible binding agents, and/or adjuvant materials can be included as

part of the composition. The tablets, pills, capsules, troches and the like can contain any of the following ingredients, or compounds of a similar nature: a binder such as microcrystalline cellulose, gum tragacanth or gelatin; an excipient such as starch or lactose, a disintegrating agent such as alginic acid, Primogel, or corn starch; a lubricant such as magnesium stearate or Sterotes; a glidant such as colloidal silicon dioxide; a sweetening agent such as sucrose or saccharin; or a flavoring agent such as peppermint, methyl salicylate, or orange flavoring.

[338] For administration by inhalation, the compounds are delivered in the form of an aerosol spray from pressured container or dispenser which contains a suitable propellant, e.g., a gas such as carbon dioxide, or a nebulizer.

[339] Systemic administration can also be by transmucosal or transdermal means. For transmucosal or transdermal administration, penetrants appropriate to the barrier to be permeated are used in the formulation. Such penetrants are generally known in the art, and include, for example, for transmucosal administration, detergents, bile salts, and fusidic acid derivatives. Transmucosal administration can be accomplished through the use of nasal sprays or suppositories. For transdermal administration, the active compounds are formulated into ointments, salves, gels, or creams as generally known in the art.

[340] The compounds can also be prepared in the form of suppositories (e.g., with conventional suppository bases such as cocoa butter and other glycerides) or retention enemas for rectal delivery.

[341] In one embodiment, the active compounds are prepared with carriers that will protect the compound against rapid elimination from the body, such as a controlled release formulation, including implants and microencapsulated delivery systems. Biodegradable, biocompatible polymers can be used, such as ethylene vinyl acetate, polyanhydrides, polyglycolic acid, collagen, polyorthoesters, and polylactic acid. Methods for preparation of such formulations will be apparent to those skilled in the art. The materials can also be obtained commercially from Alza Corporation and Nova Pharmaceuticals, Inc. Liposomal suspensions (including liposomes targeted to infected cells with monoclonal antibodies to viral antigens) can also be used as pharmaceutically acceptable carriers. These can be prepared according to methods known to those skilled in the art, for example, as described in U.S. Patent No. 4,522,811.

[342] It is advantageous to formulate oral or parenteral compositions in dosage unit form for ease of administration and uniformity of dosage. Dosage unit form as used herein

refers to physically discrete units suited as unitary dosages for the subject to be treated; each unit containing a predetermined quantity of active compound calculated to produce the desired therapeutic effect in association with the required pharmaceutical carrier.

[343] Toxicity and therapeutic efficacy of such compounds can be determined by standard pharmaceutical procedures in cell cultures or experimental animals, e.g., for determining the LD<sub>50</sub> (the dose lethal to 50% of the population) and the ED<sub>50</sub> (the dose therapeutically effective in 50% of the population). The dose ratio between toxic and therapeutic effects is the therapeutic index and it can be expressed as the ratio LD<sub>50</sub>/ED<sub>50</sub>. Compounds which exhibit high therapeutic indices are preferred. While compounds that exhibit toxic side effects may be used, care should be taken to design a delivery system that targets such compounds to the site of affected tissue in order to minimize potential damage to uninfected cells and, thereby, reduce side effects.

\* [344] The data obtained from the cell culture assays and animal studies can be used in formulating a range of dosage for use in humans. The dosage of such compounds lies preferably within a range of circulating concentrations that include the ED<sub>50</sub> with little or no toxicity. The dosage may vary within this range depending upon the dosage form employed and the route of administration utilized. For any compound used in the method of the invention, the therapeutically effective dose can be estimated initially from cell culture assays. A dose may be formulated in animal models to achieve a circulating plasma concentration range that includes the IC<sub>50</sub> (i.e., the concentration of the test compound which achieves a half-maximal inhibition of symptoms) as determined in cell culture. Such information can be used to more accurately determine useful doses in humans. Levels in plasma may be measured, for example, by high performance liquid chromatography.

[345] As defined herein, a therapeutically effective amount of protein or polypeptide (i.e., an effective dosage) ranges from about 0.001 to 30 mg/kg body weight, preferably about 0.01 to 25 mg/kg body weight, more preferably about 0.1 to 20 mg/kg body weight, and even more preferably about 1 to 10 mg/kg, 2 to 9 mg/kg, 3 to 8 mg/kg, 4 to 7 mg/kg, or 5 to 6 mg/kg body weight. The protein or polypeptide can be administered one time per week for between about 1 to 10 weeks, preferably between 2 to 8 weeks, more preferably between about 3 to 7 weeks, and even more preferably for about 4, 5, or 6 weeks. The skilled artisan will appreciate that certain factors may influence the dosage and timing required to effectively treat a subject, including but not limited to the severity of the disease or disorder, previous treatments, the general health and/or age of the subject, and other

diseases present. Moreover, treatment of a subject with a therapeutically effective amount of a protein, polypeptide, or antibody can include a single treatment or, preferably, can include a series of treatments.

[346] For antibodies, the preferred dosage is 0.1 mg/kg of body weight (generally 10 mg/kg to 20 mg/kg). If the antibody is to act in the brain, a dosage of 50 mg/kg to 100 mg/kg is usually appropriate. Generally, partially human antibodies and fully human antibodies have a longer half-life within the human body than other antibodies.

Accordingly, lower dosages and less frequent administration is often possible. Modifications such as lipidation can be used to stabilize antibodies and to enhance uptake and tissue penetration (e.g., into the brain). A method for lipidation of antibodies is described by Cruikshank *et al.* ((1997) *J. Acquired Immune Deficiency Syndromes and Human Retrovirology* 14:193).

[347] The present invention encompasses agents which modulate expression or activity. An agent may, for example, be a small molecule. For example, such small molecules include, but are not limited to, peptides, peptidomimetics (e.g., peptoids), amino acids, amino acid analogs, polynucleotides, polynucleotide analogs, nucleotides, nucleotide analogs, organic or inorganic compounds (i.e., including heteroorganic and organometallic compounds) having a molecular weight less than about 10,000 grams per mole, organic or inorganic compounds having a molecular weight less than about 5,000 grams per mole, organic or inorganic compounds having a molecular weight less than about 1,000 grams per mole, organic or inorganic compounds having a molecular weight less than about 500 grams per mole, and salts, esters, and other pharmaceutically acceptable forms of such compounds.

[348] Exemplary doses include milligram or microgram amounts of the small molecule per kilogram of subject or sample weight (e.g., about 1 microgram per kilogram to about 500 milligrams per kilogram, about 100 micrograms per kilogram to about 5 milligrams per kilogram, or about 1 microgram per kilogram to about 50 micrograms per kilogram. It is furthermore understood that appropriate doses of a small molecule depend upon the potency of the small molecule with respect to the expression or activity to be modulated. When one or more of these small molecules is to be administered to an animal (e.g., a human) in order to modulate expression or activity of a polypeptide or nucleic acid of the invention, a physician, veterinarian, or researcher may, for example, prescribe a relatively low dose at first, subsequently increasing the dose until an appropriate response is obtained. In addition, it is understood that the specific dose level for any particular animal

subject will depend upon a variety of factors including the activity of the specific compound employed, the age, body weight, general health, gender, and diet of the subject, the time of administration, the route of administration, the rate of excretion, any drug combination, and the degree of expression or activity to be modulated.

[349] An antibody (or fragment thereof) may be conjugated to a therapeutic moiety such as a cytotoxin, a therapeutic agent or a radioactive metal ion. A cytotoxin or cytotoxic agent includes any agent that is detrimental to cells. Examples include taxol, cytochalasin B, gramicidin D, ethidium bromide, emetine, mitomycin, etoposide, tenoposide, vincristine, vinblastine, colchicin, doxorubicin, daunorubicin, dihydroxy anthracin dione, mitoxantrone, mithramycin, actinomycin D, 1-dehydrotestosterone, glucocorticoids, procaine, tetracaine, lidocaine, propranolol, and puromycin and analogs or homologs thereof. Therapeutic agents include, but are not limited to, antimetabolites (e.g., methotrexate, 6-mercaptopurine, 6-thioguanine, cytarabine, 5-fluorouracil decarbazine), alkylating agents (e.g., mechlorethamine, thioepa chlorambucil, melphalan, carmustine (BSNU) and lomustine (CCNU); cyclothosphamide, busulfan, dibromomannitol, streptozotocin, mitomycin C, and cis-dichlorodiamine platinum (II) (DDP) cisplatin), anthracyclines (e.g., daunorubicin (formerly daunomycin) and doxorubicin), antibiotics (e.g., dactinomycin (formerly actinomycin), bleomycin, mithramycin, and anthramycin (AMC)), and anti-mitotic agents (e.g., vincristine and vinblastine).

[350] The conjugates of the invention can be used for modifying a given biological response, the drug moiety is not to be construed as limited to classical chemical therapeutic agents. For example, the drug moiety may be a protein or polypeptide possessing a desired biological activity. Such proteins may include, for example, a toxin such as abrin, ricin A, pseudomonas exotoxin, or diphtheria toxin; a protein such as tumor necrosis factor,  $\alpha$ -interferon,  $\beta$ -interferon, nerve growth factor, platelet derived growth factor, tissue plasminogen activator; or, biological response modifiers such as, for example, lymphokines, interleukin-1 ("IL-1"), interleukin-2 ("IL-2"), interleukin-6 ("IL-6"), granulocyte macrophase colony stimulating factor ("GM-CSF"), granulocyte colony stimulating factor ("G-CSF"), or other growth factors.

[351] Alternatively, an antibody can be conjugated to a second antibody to form an antibody heteroconjugate as described by Segal in U.S. Patent No. 4,676,980.

[352] The nucleic acid molecules of the invention can be inserted into vectors and used as gene therapy vectors. Gene therapy vectors can be delivered to a subject by, for

example, intravenous injection, local administration (see U.S. Patent 5,328,470) or by stereotactic injection (see e.g., Chen *et al.* (1994) *Proc. Natl. Acad. Sci. USA* 91:3054-3057). The pharmaceutical preparation of the gene therapy vector can include the gene therapy vector in an acceptable diluent, or can comprise a slow release matrix in which the gene delivery vehicle is imbedded. Alternatively, where the complete gene delivery vector can be produced intact from recombinant cells, e.g., retroviral vectors, the pharmaceutical preparation can include one or more cells which produce the gene delivery system.

[353] The pharmaceutical compositions can be included in a container, pack, or dispenser together with instructions for administration.

#### Methods of Treatment

[354] The present invention provides for both prophylactic and therapeutic methods of treating a subject at risk of (or susceptible to) a disorder or having a disorder associated with aberrant or unwanted 21953 expression or activity. As used herein, the term "treatment" is defined as the application or administration of a therapeutic agent to a patient, or application or administration of a therapeutic agent to an isolated tissue or cell line from a patient, who has a disease, a symptom of disease or a predisposition toward a disease, with the purpose to cure, heal, alleviate, relieve, alter, remedy, ameliorate, improve or affect the disease, the symptoms of disease or the predisposition toward disease. A therapeutic agent includes, but is not limited to, small molecules, peptides, antibodies, ribozymes and antisense oligonucleotides.

[355] With regards to both prophylactic and therapeutic methods of treatment, such treatments may be specifically tailored or modified, based on knowledge obtained from the field of pharmacogenomics. "Pharmacogenomics", as used herein, refers to the application of genomics technologies such as gene sequencing, statistical genetics, and gene expression analysis to drugs in clinical development and on the market. More specifically, the term refers the study of how a patient's genes determine his or her response to a drug (e.g., a patient's "drug response phenotype", or "drug response genotype".) Thus, another aspect of the invention provides methods for tailoring an individual's prophylactic or therapeutic treatment with either the 21953 molecules of the present invention or 21953 modulators according to that individual's drug response genotype. Pharmacogenomics allows a clinician or physician to target prophylactic or therapeutic treatments to patients who will

most benefit from the treatment and to avoid treatment of patients who will experience toxic drug-related side effects.

[356] In one aspect, the invention provides a method for preventing in a subject, a disease or condition associated with an aberrant or unwanted 21953 expression or activity, by administering to the subject a 21953 or an agent which modulates 21953 expression or at least one 21953 activity. Subjects at risk for a disease which is caused or contributed to by aberrant or unwanted 21953 expression or activity can be identified by, for example, any or a combination of diagnostic or prognostic assays as described herein. Administration of a prophylactic agent can occur prior to the manifestation of symptoms characteristic of the 21953 aberrance, such that a disease or disorder is prevented or, alternatively, delayed in its progression. Depending on the type of 21953 aberrance, for example, a 21953, 21953 agonist or 21953 antagonist agent can be used for treating the subject. The appropriate agent can be determined based on screening assays described herein.

[357] It is possible that some 21953 disorders can be caused, at least in part, by an abnormal level of gene product, or by the presence of a gene product exhibiting abnormal activity. As such, the reduction in the level and/or activity of such gene products would bring about the amelioration of disorder symptoms.

[358] The 21953 molecules can act as novel diagnostic targets and therapeutic agents for controlling one or more of cellular proliferative and/or differentiative disorders, disorders associated with bone metabolism, immune disorders, cardiovascular disorders, liver disorders, viral diseases, pain or metabolic disorders.

[359] Examples of cellular proliferative and/or differentiative disorders include cancers and proliferative disorders mentioned above. Further examples of cancers or neoplastic conditions, in addition to the ones described above include, but are not limited to, a fibrosarcoma, myosarcoma, liposarcoma, chondrosarcoma, osteogenic sarcoma, chordoma, angiosarcoma, endothelirosarcoma, lymphangiosarcoma, lymphangioendothelirosarcoma, synovioma, mesothelioma, Ewing's tumor, leiomyosarcoma, rhabdomyosarcoma, gastric cancer, esophageal cancer, rectal cancer, pancreatic cancer, ovarian cancer, prostate cancer, uterine cancer, cancer of the head and neck, skin cancer, brain cancer, squamous cell carcinoma, sebaceous gland carcinoma, papillary carcinoma, papillary adenocarcinoma, cystadenocarcinoma, medullary carcinoma, bronchogenic carcinoma, renal cell carcinoma, hepatoma, bile duct carcinoma, choriocarcinoma, seminoma, embryonal carcinoma, Wilm's

tumor, cervical cancer, testicular cancer, small cell lung carcinoma, non-small cell lung carcinoma, bladder carcinoma, epithelial carcinoma, glioma, astrocytoma, medulloblastoma, craniopharyngioma, ependymoma, pinealoma, hemangioblastoma, acoustic neuroma, oligodendrogioma, meningioma, melanoma, neuroblastoma, retinoblastoma, leukemia, lymphoma, or Kaposi sarcoma.

[360] Disorders involving T-cells include, but are not limited to, cell-mediated hypersensitivity, such as delayed type hypersensitivity and T-cell-mediated cytotoxicity, and transplant rejection; autoimmune diseases, such as systemic lupus erythematosus, Sjögren syndrome, systemic sclerosis, inflammatory myopathies, mixed connective tissue disease, and polyarteritis nodosa and other vasculitides; immunologic deficiency syndromes, including but not limited to, primary immunodeficiencies, such as thymic hypoplasia, severe combined immunodeficiency diseases, and AIDS; leukopenia; reactive (inflammatory) proliferations of white cells, including but not limited to, leukocytosis, acute nonspecific lymphadenitis, and chronic nonspecific lymphadenitis; neoplastic proliferations of white cells, including but not limited to lymphoid neoplasms, such as precursor T-cell neoplasms, such as acute lymphoblastic leukemia/lymphoma, peripheral T-cell and natural killer cell neoplasms that include peripheral T-cell lymphoma, unspecified, adult T-cell leukemia/lymphoma, mycosis fungoides and Sézary syndrome, and Hodgkin disease.

[361] Additionally, 21953 may play an important role in the regulation of metabolism or pain disorders, e.g., by processing neuropeptides and metabolic peptide hormones. Diseases of metabolic imbalance include, but are not limited to, obesity, anorexia nervosa, cachexia, lipid disorders, and diabetes. Examples of pain disorders include, but are not limited to, pain response elicited during various forms of tissue injury, e.g., inflammation, infection, and ischemia, usually referred to as hyperalgesia (described in, for example, Fields, H.L. (1987) *Pain*, New York:McGraw-Hill); pain associated with musculoskeletal disorders, e.g., joint pain; tooth pain; headaches; pain associated with surgery; pain related to irritable bowel syndrome; or chest pain.

[362] As discussed, successful treatment of 21953 disorders can be brought about by techniques that serve to inhibit the expression or activity of target gene products. For example, compounds, e.g., an agent identified using an assays described above, that proves to exhibit negative modulatory activity, can be used in accordance with the invention to prevent and/or ameliorate symptoms of 21953 disorders. Such molecules can include, but are not limited to peptides, phosphopeptides, small organic or inorganic molecules, or

antibodies (including, for example, polyclonal, monoclonal, humanized, anti-idiotypic, chimeric or single chain antibodies, and Fab, F(ab')<sub>2</sub> and Fab expression library fragments, scFV molecules, and epitope-binding fragments thereof).

[363] Further, antisense and ribozyme molecules that inhibit expression of the target gene can also be used in accordance with the invention to reduce the level of target gene expression, thus effectively reducing the level of target gene activity. Still further, triple helix molecules can be utilized in reducing the level of target gene activity. Antisense, ribozyme and triple helix molecules are discussed above.

[364] It is possible that the use of antisense, ribozyme, and/or triple helix molecules to reduce or inhibit mutant gene expression can also reduce or inhibit the transcription (triple helix) and/or translation (antisense, ribozyme) of mRNA produced by normal target gene alleles, such that the concentration of normal target gene product present can be lower than is necessary for a normal phenotype. In such cases, nucleic acid molecules that encode and express target gene polypeptides exhibiting normal target gene activity can be introduced into cells via gene therapy method. Alternatively, in instances in that the target gene encodes an extracellular protein, it can be preferable to co-administer normal target gene protein into the cell or tissue in order to maintain the requisite level of cellular or tissue target gene activity.

[365] Another method by which nucleic acid molecules may be utilized in treating or preventing a disease characterized by 21953 expression is through the use of aptamer molecules specific for 21953 protein. Aptamers are nucleic acid molecules having a tertiary structure which permits them to specifically bind to protein ligands (see, e.g., Osborne, *et al.* (1997) *Curr. Opin. Chem Biol.* 1: 5-9; and Patel, D.J. (1997) *Curr Opin Chem Biol* 1:32-46). Since nucleic acid molecules may in many cases be more conveniently introduced into target cells than therapeutic protein molecules may be, aptamers offer a method by which 21953 protein activity may be specifically decreased without the introduction of drugs or other molecules which may have pluripotent effects.

[366] Antibodies can be generated that are both specific for target gene product and that reduce target gene product activity. Such antibodies may, therefore, be administered in instances whereby negative modulatory techniques are appropriate for the treatment of 21953 disorders. For a description of antibodies, see the Antibody section above.

[367] In circumstances wherein injection of an animal or a human subject with a 21953 protein or epitope for stimulating antibody production is harmful to the subject, it is

possible to generate an immune response against 21953 through the use of anti-idiotypic antibodies (see, for example, Herlyn, D. (1999) *Ann Med* 31:66-78; and Bhattacharya-Chatterjee, M., and Foon, K.A. (1998) *Cancer Treat Res.* 94:51-68). If an anti-idiotypic antibody is introduced into a mammal or human subject, it should stimulate the production of anti-anti-idiotypic antibodies, which should be specific to the 21953 protein. Vaccines directed to a disease characterized by 21953 expression may also be generated in this fashion.

[368] In instances where the target antigen is intracellular and whole antibodies are used, internalizing antibodies may be preferred. Lipofectin or liposomes can be used to deliver the antibody or a fragment of the Fab region that binds to the target antigen into cells. Where fragments of the antibody are used, the smallest inhibitory fragment that binds to the target antigen is preferred. For example, peptides having an amino acid sequence corresponding to the Fv region of the antibody can be used. Alternatively, single chain neutralizing antibodies that bind to intracellular target antigens can also be administered. Such single chain antibodies can be administered, for example, by expressing nucleotide sequences encoding single-chain antibodies within the target cell population (see e.g., Marasco *et al.* (1993) *Proc. Natl. Acad. Sci. USA* 90:7889-7893).

[369] The identified compounds that inhibit target gene expression, synthesis and/or activity can be administered to a patient at therapeutically effective doses to prevent, treat or ameliorate 21953 disorders. A therapeutically effective dose refers to that amount of the compound sufficient to result in amelioration of symptoms of the disorders. Toxicity and therapeutic efficacy of such compounds can be determined by standard pharmaceutical procedures as described above.

[370] The data obtained from the cell culture assays and animal studies can be used in formulating a range of dosage for use in humans. The dosage of such compounds lies preferably within a range of circulating concentrations that include the ED<sub>50</sub> with little or no toxicity. The dosage can vary within this range depending upon the dosage form employed and the route of administration utilized. For any compound used in the method of the invention, the therapeutically effective dose can be estimated initially from cell culture assays. A dose can be formulated in animal models to achieve a circulating plasma concentration range that includes the IC<sub>50</sub> (i.e., the concentration of the test compound that achieves a half-maximal inhibition of symptoms) as determined in cell culture. Such

information can be used to more accurately determine useful doses in humans. Levels in plasma can be measured, for example, by high performance liquid chromatography.

[371] Another example of determination of effective dose for an individual is the ability to directly assay levels of "free" and "bound" compound in the serum of the test subject. Such assays may utilize antibody mimics and/or "biosensors" that have been created through molecular imprinting techniques. The compound which is able to modulate 21953 activity is used as a template, or "imprinting molecule", to spatially organize polymerizable monomers prior to their polymerization with catalytic reagents. The subsequent removal of the imprinted molecule leaves a polymer matrix which contains a repeated "negative image" of the compound and is able to selectively rebind the molecule under biological assay conditions. A detailed review of this technique can be seen in Ansell, R. J. *et al* (1996) *Current Opinion in Biotechnology* 7:89-94 and in Shea, K.J. (1994) *Trends in Polymer Science* 2:166-173. Such "imprinted" affinity matrixes are amenable to ligand-binding assays, whereby the immobilized monoclonal antibody component is replaced by an appropriately imprinted matrix. An example of the use of such matrixes in this way can be seen in Vlatakis, G. *et al* (1993) *Nature* 361:645-647. Through the use of isotope-labeling, the "free" concentration of compound which modulates the expression or activity of 21953 can be readily monitored and used in calculations of IC<sub>50</sub>.

[372] Such "imprinted" affinity matrixes can also be designed to include fluorescent groups whose photon-emitting properties measurably change upon local and selective binding of target compound. These changes can be readily assayed in real time using appropriate fiberoptic devices, in turn allowing the dose in a test subject to be quickly optimized based on its individual IC<sub>50</sub>. An rudimentary example of such a "biosensor" is discussed in Kriz, D. *et al* (1995) *Analytical Chemistry* 67:2142-2144.

[373] Another aspect of the invention pertains to methods of modulating 21953 expression or activity for therapeutic purposes. Accordingly, in an exemplary embodiment, the modulatory method of the invention involves contacting a cell with a 21953 or agent that modulates one or more of the activities of 21953 protein activity associated with the cell. An agent that modulates 21953 protein activity can be an agent as described herein, such as a nucleic acid or a protein, a naturally-occurring target molecule of a 21953 protein (e.g., a 21953 substrate or receptor), a 21953 antibody, a 21953 agonist or antagonist, a peptidomimetic of a 21953 agonist or antagonist, or other small molecule.

[374] In one embodiment, the agent stimulates one or 21953 activities. Examples of such stimulatory agents include active 21953 protein and a nucleic acid molecule encoding 21953. In another embodiment, the agent inhibits one or more 21953 activities. Examples of such inhibitory agents include antisense 21953 nucleic acid molecules, anti-21953 antibodies, and 21953 inhibitors. These modulatory methods can be performed *in vitro* (e.g., by culturing the cell with the agent) or, alternatively, *in vivo* (e.g., by administering the agent to a subject). As such, the present invention provides methods of treating an individual afflicted with a disease or disorder characterized by aberrant or unwanted expression or activity of a 21953 protein or nucleic acid molecule. In one embodiment, the method involves administering an agent (e.g., an agent identified by a screening assay described herein), or combination of agents that modulates (e.g., up regulates or down regulates) 21953 expression or activity. In another embodiment, the method involves administering a 21953 protein or nucleic acid molecule as therapy to compensate for reduced, aberrant, or unwanted 21953 expression or activity.

[375] Stimulation of 21953 activity is desirable in situations in which 21953 is abnormally downregulated and/or in which increased 21953 activity is likely to have a beneficial effect. For example, stimulation of 21953 activity is desirable in situations in which a 21953 is downregulated and/or in which increased 21953 activity is likely to have a beneficial effect. Likewise, inhibition of 21953 activity is desirable in situations in which 21953 is abnormally upregulated and/or in which decreased 21953 activity is likely to have a beneficial effect.

### Pharmacogenomics

[376] The 21953 molecules of the present invention, as well as agents, or modulators which have a stimulatory or inhibitory effect on 21953 activity (e.g., 21953 gene expression) as identified by a screening assay described herein can be administered to individuals to treat (prophylactically or therapeutically) 21953 associated disorders (e.g., a cell proliferative or cell differentiative disorder, e.g., a cancer, e.g., a cancer of the lung, prostate, breast, or colon) associated with aberrant or unwanted 21953 activity. In conjunction with such treatment, pharmacogenomics (i.e., the study of the relationship between an individual's genotype and that individual's response to a foreign compound or drug) may be considered. Differences in metabolism of therapeutics can lead to severe toxicity or therapeutic failure by altering the relation between dose and blood concentration of the

pharmacologically active drug. Thus, a physician or clinician may consider applying knowledge obtained in relevant pharmacogenomics studies in determining whether to administer a 21953 molecule or 21953 modulator as well as tailoring the dosage and/or therapeutic regimen of treatment with a 21953 molecule or 21953 modulator.

[377] Pharmacogenomics deals with clinically significant hereditary variations in the response to drugs due to altered drug disposition and abnormal action in affected persons. See, for example, Eichelbaum, M. *et al.* (1996) *Clin. Exp. Pharmacol. Physiol.* 23:983-985 and Linder, M.W. *et al.* (1997) *Clin. Chem.* 43:254-266. In general, two types of pharmacogenetic conditions can be differentiated. Genetic conditions transmitted as a single factor altering the way drugs act on the body (altered drug action) or genetic conditions transmitted as single factors altering the way the body acts on drugs (altered drug metabolism). These pharmacogenetic conditions can occur either as rare genetic defects or as naturally-occurring polymorphisms. For example, glucose-6-phosphate dehydrogenase deficiency (G6PD) is a common inherited enzymopathy in which the main clinical complication is haemolysis after ingestion of oxidant drugs (anti-malarials, sulfonamides, analgesics, nitrofurans) and consumption of fava beans.

[378] One pharmacogenomics approach to identifying genes that predict drug response, known as "a genome-wide association", relies primarily on a high-resolution map of the human genome consisting of already known gene-related markers (e.g., a "bi-allelic" gene marker map which consists of 60,000-100,000 polymorphic or variable sites on the human genome, each of which has two variants.) Such a high-resolution genetic map can be compared to a map of the genome of each of a statistically significant number of patients taking part in a Phase II/III drug trial to identify markers associated with a particular observed drug response or side effect. Alternatively, such a high resolution map can be generated from a combination of some ten-million known single nucleotide polymorphisms (SNPs) in the human genome. As used herein, a "SNP" is a common alteration that occurs in a single nucleotide base in a stretch of DNA. For example, a SNP may occur once per every 1000 bases of DNA. A SNP may be involved in a disease process, however, the vast majority may not be disease-associated. Given a genetic map based on the occurrence of such SNPs, individuals can be grouped into genetic categories depending on a particular pattern of SNPs in their individual genome. In such a manner, treatment regimens can be tailored to groups of genetically similar individuals, taking into account traits that may be common among such genetically similar individuals.

[379] Alternatively, a method termed the “candidate gene approach,” can be utilized to identify genes that predict drug response. According to this method, if a gene that encodes a drug’s target is known (e.g., a 21953 protein of the present invention), all common variants of that gene can be fairly easily identified in the population and it can be determined if having one version of the gene versus another is associated with a particular drug response.

[380] Alternatively, a method termed the “gene expression profiling,” can be utilized to identify genes that predict drug response. For example, the gene expression of an animal dosed with a drug (e.g., a 21953 molecule or 21953 modulator of the present invention) can give an indication whether gene pathways related to toxicity have been turned on.

[381] Information generated from more than one of the above pharmacogenomics approaches can be used to determine appropriate dosage and treatment regimens for prophylactic or therapeutic treatment of an individual. This knowledge, when applied to dosing or drug selection, can avoid adverse reactions or therapeutic failure and thus enhance therapeutic or prophylactic efficiency when treating a subject with a 21953 molecule or 21953 modulator, such as a modulator identified by one of the exemplary screening assays described herein.

[382] The present invention further provides methods for identifying new agents, or combinations, that are based on identifying agents that modulate the activity of one or more of the gene products encoded by one or more of the 21953 genes of the present invention, wherein these products may be associated with resistance of the cells to a therapeutic agent. Specifically, the activity of the proteins encoded by the 21953 genes of the present invention can be used as a basis for identifying agents for overcoming agent resistance. By blocking the activity of one or more of the resistance proteins, target cells, e.g., human cells, will become sensitive to treatment with an agent that the unmodified target cells were resistant to.

[383] Monitoring the influence of agents (e.g., drugs) on the expression or activity of a 21953 protein can be applied in clinical trials. For example, the effectiveness of an agent determined by a screening assay as described herein to increase 21953 gene expression, protein levels, or upregulate 21953 activity, can be monitored in clinical trials of subjects exhibiting decreased 21953 gene expression, protein levels, or downregulated 21953 activity. Alternatively, the effectiveness of an agent determined by a screening assay to decrease 21953 gene expression, protein levels, or downregulate 21953 activity, can be monitored in clinical trials of subjects exhibiting increased 21953 gene expression, protein

levels, or upregulated 21953 activity. In such clinical trials, the expression or activity of a 21953 gene, and preferably, other genes that have been implicated in, for example, a 21953-associated disorder can be used as a “read out” or markers of the phenotype of a particular cell.

### **21953 Informatics**

[384] The sequence of a 21953 molecule is provided in a variety of media to facilitate use thereof. A sequence can be provided as a manufacture, other than an isolated nucleic acid or amino acid molecule, which contains a 21953. Such a manufacture can provide a nucleotide or amino acid sequence, e.g., an open reading frame, in a form which allows examination of the manufacture using means not directly applicable to examining the nucleotide or amino acid sequences, or a subset thereof, as they exists in nature or in purified form. The sequence information can include, but is not limited to, 21953 full-length nucleotide and/or amino acid sequences, partial nucleotide and/or amino acid sequences, polymorphic sequences including single nucleotide polymorphisms (SNPs), epitope sequence, and the like. In a preferred embodiment, the manufacture is a machine-readable medium, e.g., a magnetic, optical, chemical or mechanical information storage device.

[385] As used herein, “machine-readable media” refers to any medium that can be read and accessed directly by a machine, e.g., a digital computer or analogue computer. Non-limiting examples of a computer include a desktop PC, laptop, mainframe, server (e.g., a web server, network server, or server farm), handheld digital assistant, pager, mobile telephone, and the like. The computer can be stand-alone or connected to a communications network, e.g., a local area network (such as a VPN or intranet), a wide area network (e.g., an Extranet or the Internet), or a telephone network (e.g., a wireless, DSL, or ISDN network). Machine-readable media include, but are not limited to: magnetic storage media, such as floppy discs, hard disc storage medium, and magnetic tape; optical storage media such as CD-ROM; electrical storage media such as RAM, ROM, EPROM, EEPROM, flash memory, and the like; and hybrids of these categories such as magnetic/optical storage media.

[386] A variety of data storage structures are available to a skilled artisan for creating a machine-readable medium having recorded thereon a nucleotide or amino acid sequence of the present invention. The choice of the data storage structure will generally be based on the means chosen to access the stored information. In addition, a variety of data processor

programs and formats can be used to store the nucleotide sequence information of the present invention on computer readable medium. The sequence information can be represented in a word processing text file, formatted in commercially-available software such as WordPerfect and Microsoft Word, or represented in the form of an ASCII file, stored in a database application, such as DB2, Sybase, Oracle, or the like. The skilled artisan can readily adapt any number of data processor structuring formats (e.g., text file or database) in order to obtain computer readable medium having recorded thereon the nucleotide sequence information of the present invention.

[387] In a preferred embodiment, the sequence information is stored in a relational database (such as Sybase or Oracle). The database can have a first table for storing sequence (nucleic acid and/or amino acid sequence) information. The sequence information can be stored in one field (e.g., a first column) of a table row and an identifier for the sequence can be stored in another field (e.g., a second column) of the table row. The database can have a second table, e.g., storing annotations. The second table can have a field for the sequence identifier, a field for a descriptor or annotation text (e.g., the descriptor can refer to a functionality of the sequence, a field for the initial position in the sequence to which the annotation refers, and a field for the ultimate position in the sequence to which the annotation refers. Non-limiting examples for annotation to nucleic acid sequences include polymorphisms (e.g., SNP's) translational regulatory sites and splice junctions. Non-limiting examples for annotations to amino acid sequence include polypeptide domains, e.g., a domain described herein; active sites and other functional amino acids; and modification sites.

[388] By providing the nucleotide or amino acid sequences of the invention in computer readable form, the skilled artisan can routinely access the sequence information for a variety of purposes. For example, one skilled in the art can use the nucleotide or amino acid sequences of the invention in computer readable form to compare a target sequence or target structural motif with the sequence information stored within the data storage means. A search is used to identify fragments or regions of the sequences of the invention which match a particular target sequence or target motif. The search can be a BLAST search or other routine sequence comparison, e.g., a search described herein.

[389] Thus, in one aspect, the invention features a method of analyzing 21953, e.g., analyzing structure, function, or relatedness to one or more other nucleic acid or amino acid sequences. The method includes: providing a 21953 nucleic acid or amino acid sequence;

comparing the 21953 sequence with a second sequence, e.g., one or more preferably a plurality of sequences from a collection of sequences, e.g., a nucleic acid or protein sequence database to thereby analyze 21953. The method can be performed in a machine, e.g., a computer, or manually by a skilled artisan.

[390] The method can include evaluating the sequence identity between a 21953 sequence and a database sequence. The method can be performed by accessing the database at a second site, e.g., over the Internet.

[391] As used herein, a "target sequence" can be any DNA or amino acid sequence of six or more nucleotides or two or more amino acids. A skilled artisan can readily recognize that the longer a target sequence is, the less likely a target sequence will be present as a random occurrence in the database. Typical sequence lengths of a target sequence are from about 10 to 100 amino acids or from about 30 to 300 nucleotide residues. However, it is well recognized that commercially important fragments, such as sequence fragments involved in gene expression and protein processing, may be of shorter length.

[392] Computer software is publicly available which allows a skilled artisan to access sequence information provided in a computer readable medium for analysis and comparison to other sequences. A variety of known algorithms are disclosed publicly and a variety of commercially available software for conducting search means are and can be used in the computer-based systems of the present invention. Examples of such software include, but are not limited to, MacPattern (EMBL), BLASTN and BLASTX (NCBI).

[393] Thus, the invention features a method of making a computer readable record of a sequence of a 21953 sequence which includes recording the sequence on a computer readable matrix. In a preferred embodiment the record includes one or more of the following: identification of an ORF; identification of a domain, region, or site; identification of the start of transcription; identification of the transcription terminator; the full length amino acid sequence of the protein, or a mature form thereof; the 5' end of the translated region.

[394] In another aspect, the invention features, a method of analyzing a sequence. The method includes: providing a 21953 sequence, or record, in machine-readable form; comparing a second sequence to the 21953 sequence; thereby analyzing a sequence. Comparison can include comparing to sequences for sequence identity or determining if one sequence is included within the other, e.g., determining if the 21953 sequence includes a sequence being compared. In a preferred embodiment the 21953 or second sequence is

stored on a first computer, e.g., at a first site and the comparison is performed, read, or recorded on a second computer, e.g., at a second site. For example, the 21953 or second sequence can be stored in a public or proprietary database in one computer, and the results of the comparison performed, read, or recorded on a second computer. In a preferred embodiment the record includes one or more of the following: identification of an ORF; identification of a domain, region, or site; identification of the start of transcription; identification of the transcription terminator; the full length amino acid sequence of the protein, or a mature form thereof; the 5' end of the translated region.

[395] In another aspect, the invention provides a machine-readable medium for holding instructions for performing a method for determining whether a subject has a 21953-associated disease or disorder or a pre-disposition to a 21953-associated disease or disorder, wherein the method comprises the steps of determining 21953 sequence information associated with the subject and based on the 21953 sequence information, determining whether the subject has a 21953-associated disease or disorder or a pre-disposition to a 21953-associated disease or disorder and/or recommending a particular treatment for the disease, disorder or pre-disease condition.

[396] The invention further provides in an electronic system and/or in a network, a method for determining whether a subject has a 21953-associated disease or disorder or a pre-disposition to a disease associated with a 21953 wherein the method comprises the steps of determining 21953 sequence information associated with the subject, and based on the 21953 sequence information, determining whether the subject has a 21953-associated disease or disorder or a pre-disposition to a 21953-associated disease or disorder, and/or recommending a particular treatment for the disease, disorder or pre-disease condition. In a preferred embodiment, the method further includes the step of receiving information, e.g., phenotypic or genotypic information, associated with the subject and/or acquiring from a network phenotypic information associated with the subject. The information can be stored in a database, e.g., a relational database. In another embodiment, the method further includes accessing the database, e.g., for records relating to other subjects, comparing the 21953 sequence of the subject to the 21953 sequences in the database to thereby determine whether the subject has a 21953-associated disease or disorder, or a pre-disposition for such.

[397] The present invention also provides in a network, a method for determining whether a subject has a 21953 associated disease or disorder or a pre-disposition to a 21953-associated disease or disorder associated with 21953, said method comprising the steps of

receiving 21953 sequence information from the subject and/or information related thereto, receiving phenotypic information associated with the subject, acquiring information from the network corresponding to 21953 and/or corresponding to a 21953-associated disease or disorder (e.g., a cell proliferative or cell differentiative disorder, e.g., a cancer, e.g., a cancer of the lung, prostate, breast, or colon), and based on one or more of the phenotypic information, the 21953 information (e.g., sequence information and/or information related thereto), and the acquired information, determining whether the subject has a 21953-associated disease or disorder or a pre-disposition to a 21953-associated disease or disorder. The method may further comprise the step of recommending a particular treatment for the disease, disorder or pre-disease condition.

[398] The present invention also provides a method for determining whether a subject has a 21953 -associated disease or disorder or a pre-disposition to a 21953-associated disease or disorder, said method comprising the steps of receiving information related to 21953 (e.g., sequence information and/or information related thereto), receiving phenotypic information associated with the subject, acquiring information from the network related to 21953 and/or related to a 21953-associated disease or disorder, and based on one or more of the phenotypic information, the 21953 information, and the acquired information, determining whether the subject has a 21953-associated disease or disorder or a pre-disposition to a 21953-associated disease or disorder. The method may further comprise the step of recommending a particular treatment for the disease, disorder or pre-disease condition.

[399] This invention is further illustrated by the following examples that should not be construed as limiting. The contents of all references, patents and published patent applications cited throughout this application are incorporated herein by reference.

### EXAMPLES

#### Example 1: Identification and Characterization of Human 21953 cDNA

[400] The human 21953 nucleic acid sequence is recited as follows:

[401] CTATAGGGAGTCGCCACGCGTCCGGCCTCCGAGGCCAAGGCCGCTG  
CTACTGCCGCCGCTGCTTCTTAGTGCCGCGITCGCCGCCTGGGTGTCAACCGGCG  
CCGCCGCCGAGGAAGCCACTACAACCAGGACCGGAGTGGAGGCCGGCAGCAT

GAAGCGGCGCAGGCCGCTCCATAGCGCACGTGGACGGTCCGGGCGGGCC  
GGGGGAAGGAAAATGCAACATGGCAGCAGCAATGGAAACAGAACAGCTGG  
TGTTGAGATATTGAAACTCGGACTGTGAGGAGAATATTGAATCACAGGATCG  
GCCTAAATTGGAGCCCTTTATGTGAGCGGTATTCTGGAGTCAGCTAAAAA  
GCTGCTGCCGATACCAGAAAATCATGGCTACATGATGGCTAAGGCACCACA  
TGATTTCATGTTGTGAAGAGGAATGATCCAGATGGACCTCATTAGACAGAAC  
CTATTACCTGCCATGTCGGTGAGAACAGAGAAAATACACTGTTTATTCTGAA  
ATTCCAAAACATCAATAGAGCAGCAGTCTTAATGCTCTTGGAGCCTCTT  
TGGATCTTTTCAAGCAACACTGGACTATGGAATGTATTCTGAGAACAGAAC  
TATTAAGAGAAAGAAAACGATTGGAACAGTCGGATTGCTTACGATTATC  
ACCAAGGAAGTGGAACATTCTGTTCAAGCCGGTAGTGGATTATCACGTAA  
AAGATGGAGGCCACAAGGATTACGCAACACCTTAAGGCCAATCTAGTG  
GAAACTAGTTGCCAACATACGGATGGATCCAAAATTATGCCCTGCTGATCCA  
GACTGGATTGCTTTATACATAGCAACGATAATTGGATATCTAACATCGTAACCA  
GAGAAGAAAGGAGACTCACTTATGTGCACAATGAGCTAGCCAACATGGAAGAA  
GATGCCAGATCAGCTGGAGTCGCTACCTTGTCTCCAAGAAGAACATTGATAGA  
TATTCTGGCTATTGGTGGTCCAAAAGCTGAAACAACCTCCAGTGGTGGTAAA  
ATTCTAGAATTCTATATGAAGAAAATGATGAATCTGAGGTGGAAATTATTCTAT  
GTTACATCCCCTATGTGGAAACAAGGAGGGCAGATTCAATTCCGTTATCCTAAA  
ACAGGTACAGCAAATCTAAAGTCACCTTAAGATGTCAGAAATAATGATTGAT  
GCTGAAGGAAGGATCATAGATGTCAGATAAGGAACAAATTCAACCTTTGAG  
ATTCTATTGAAGGAGTTGAATATATTGCCAGAGCTGGATGGACTCCTGAGGGA  
AAATATGCTTGGTCCATCCTACTAGATCGCTCCAGACTCGCCTGCAGATAGTGT  
TGATCTCACCTGAATTATTATCCAGTAGAAGATGATGTTATGGAAAGGCAGA  
GACTCATTGAGTCAGTGCCTGATTCTGTGACGCCACTAATTATCTATGAAGAAA  
CAACAGACATCTGGATAAAATCCATGACATCTTCATGTTTCCCAAAGTC  
CGAAGAGGAAATTGAGTTATTGGCCTCTGAATGCAAAACAGGTTCCGTCA  
TTTATACAAAATTACATCTATTAAAGGAAAGCAAATATAAACGATCCAGTGG  
TGGGCTGCCTGCTCCAAGTGATTCAAGTGTCTATCAAAGAGGAGATAGCAAT  
TACCAAGTGGTGAATGGGAAGTTCTGGCCGGCATGGATCTAATATCCAAGTGA  
TGAAGTCAGAAGGCTGGTATATTGAAGGCACCAAAGACTCCCTTAGAGCA  
TCACCTGTACGTAGTCAGTACGTAATCCTGGAGAGGTGACAAGGCTGACTGA  
CCGTGGCTACTCACATTCTGCTGCATCAGTCAGCACTGTGACTCTTATAAGT

AAGTATAGTAACCAGAAGAACCCACACTGTGTGCCCTTACAAGCTATCAAGT  
 CCTGAAGATGACCCAACCTGCAAAACAAAGGAATTGGGCCACCATTGGAT  
 TCAGCAGGTCCCTTCCTGACTATACTCCTCCAGAAATTCTCTTTGAAAGTA  
 CTACTGGATTACATTGTATGGGATGCTCTACAAGCCTCATGATCTACAGCCTGG  
 AAAGAAAATATCCTACTGTGCTGTCATATGGTGGCCTCAGGTGCAGTTGGT  
 GAATAATCGGTTAAAGGAGTCAAGTATTCCGCTTGAATACCCTAGCCTCTCTA  
 GGTATGTGGTTGAGTGTAGAGACAACAGGGGATCCTGTCACCGAGGGCTTAAA  
 TTTGAAGGCGCCTTAAATATAAAATGGGTCAAATAGAAATTGACGATCAGGTG  
 GAAGGACTCCAATATCTAGCTTCTCGATATGATTCAATTGACTTAGATCGTGTGG  
 GCATCCACGGCTGGCCTATGGAGGATACCTCTCCCTGATGGCATTATGCAGA  
 GGTCAAGATATCTCAGGGTTGCTATTGCTGGGCCCCAGTCACTCTGTGGATCTT  
 CTATGATAACAGGATAACCGAACGTTATATGGGTACCCCTGACCAGAACATGAACA  
~~GGGCTATTACTTAGGATCTGTGGCCATGCAAGCAGAAAAGTCCCCTCTGAACC~~  
~~AAATCGTTACTGCTTACATGGTTCTGGATGAGAATGTCCATTITGCACAT~~  
~~ACCACTATATTACTGAGTTTTAGTGAGGGCTGGAAAGCCATATGATTACAG~~  
~~ATCTATCCTCAGGAGAGACACAGCATAAGAGTCTGAAATCGGGAGAACATTAT~~  
~~GAACTCATCTTGCACACCTCAAGAAAACCTGGATCACGTATTGCTGCTC~~  
~~TAAAAGTGATAATTTGACCTGTGTAGAACTCTCTGGTATACACTGGCTATT~~  
~~AAACAAATGAGGAGGTTAACACAGAAAACACAGAACATTGATCATCACATT~~  
~~GATACCTGCCATGTAACATCTACTCCTGAAAATAATGTGGTGCCTGAGGGG~~  
~~TCTACGGTTGTGGTAGTAATCTAACCTAACCCACATGCTCAAATCAAAT~~  
~~GATACATATTCCCTGAGAGACCCAGCAATACCATAAGAATTACTAAAAAAA~~  
~~AAAAAAA ( SEQ ID NO:1).~~

[402] The human 21953 nucleic acid sequence (SEQ ID NO:1) is approximately 3143 nucleotides long. The nucleic acid sequence includes an initiation codon (ATG) and a termination codon (TAA) which are underlined above. The region between and inclusive of the initiation codon and the termination codon is a methionine-initiated coding sequence of about 2646 nucleotides (nucleotides 229-2874 of SEQ ID NO:1, designated as SEQ ID NO:3). The coding sequence encodes an 882 amino acid protein, the sequence of which is recited as follows:

[403] MAAAMETEQLGVEIFETADCEENIESQDRPKLEPFYVERYSQLKKLL  
 ADTRKYHGYMMAKAPHDFMFVKRNDPDGPHSDRIYYLAMSGENRENTLFYSEIPK  
 TINRAAVLMLSWKPLLDLFQATLDYGMYSREEELLRERKRIGTVGIASYDYHQGSG

TFLFQAGSGIYHVKGDPQGFTQQPLRPNLVETSCPNIKMDPKLCPADPDWIAFIHS  
NDIWISNIVTREERRLTYVHNELANMEEADARSAGVATFVLQEEFDYSGYWVWCPK  
AETTPSGGKILRILYEENDESEVEIIHVTSPMLETRRADSFYRYPKTGTANPKVTFKMS  
EIMIDAEGRIIDVIDKELIQPFEILFEGVEYIARAGWTPEGKYAWSILLDRSQTRLQIVL  
ISPELFIPVEDDVMERQRRIESVPDSVTPLIYEEETTDIWINIHDIFHVFQPSHEEEIEFIF  
ASECKTGFRHLYKITSILKESKYKRSSGLPAPSDFKCPKEELAITSGEWEVLGRHGS  
NIQVDEVRRLVYFEGTKDSPLEHHLYVVSYVNPGEVTRLTDGYSHSCCISQHCDFF  
ISKYSNQKNPHCVSLYKLSSPEDDPTCKTKEFWATILDSAGPLPDYTPPEIFSFESTTG  
FTLYGMLYKPHDLQPGKKYPTVLFYGGPQVQLVNNRFKGVKYFRLNTLASLGYV  
VVVIDNRGSCHRGLKFEGAFKYKMGQIEIDDQVEGLQYLASRYDFIDLDRVGIHGWS  
SYGGYLSLMAJMQRSDIFRVAIAGAPVTLWIFYDTGYTERYMGHPDQNEQGYYLG  
SVAMQAEKFPSEPNRLLLHGFLDENVHFAHTSILLSFLVRAGKPYDLQIYPQERHSI  
RVPESGEHEYELHLLHYLQENLGSRIAALKVI (SEQ ID NO:2).

**Example 2: 21953 mRNA Expression**

[404] Endogenous human 21953 gene expression was determined using the Perkin-Elmer/ABI 7700 Sequence Detection System which employs TaqMan technology. Briefly, TaqMan technology relies on standard RT-PCR with the addition of a third gene-specific oligonucleotide (referred to as a probe) which has a fluorescent dye coupled to its 5' end (typically 6-FAM) and a quenching dye at the 3' end (typically TAMRA). When the fluorescently tagged oligonucleotide is intact, the fluorescent signal from the 5' dye is quenched. As PCR proceeds, the 5' to 3' nucleolytic activity of Taq polymerase digests the labeled primer, producing a free nucleotide labeled with 6-FAM, which is now detected as a fluorescent signal. The PCR cycle where fluorescence is first released and detected is directly proportional to the starting amount of the gene of interest in the test sample, thus providing a quantitative measure of the initial template concentration. Samples were internally controlled by the addition of a second set of primers/probe specific for a reference gene such as  $\beta$ 2-macroglobulin, GAPDH which has been labeled with a different fluorophore on the 5' end (typically VIC).

[405] To determine the level of 21953 in various human tissues a primer/probe set was designed. Total RNA was prepared from a series of human tissues using an RNeasy kit from Qiagen. First strand cDNA was prepared from 1  $\mu$ g total RNA using an oligo-dT primer and Superscript II reverse transcriptase (Gibco/BRL). cDNA obtained from

approximately 50 ng total RNA was used per TaqMan reaction. Tissues tested include the human tissues and several cell lines shown in the left column of the tables below.

[406] 21953 mRNA expression was elevated in 85% of clinical lung tumor samples tested, and is similarly elevated in a number of breast tumor and colon tumor samples (see, e.g., Table 1 below).

Table 1.

Sample	Relative Expression
Breast Normal	0.02
Breast Normal	0.07
Breast Tumor	0.08
Breast Tumor	0.07
Breast Tumor	0.19
Breast Tumor	0.21
Breast Tumor	0.07
Breast Tumor	0.30
Ovary Normal	0.37
Ovary Normal	0.26
Ovary Normal	0.33
Ovary Tumor	0.16
Ovary Tumor	0.13
Ovary Tumor	0.17
Ovary Tumor	0.10
Ovary Tumor	0.12
Ovary Tumor	0.08
Ovary Tumor	0.52
Ovary Tumor	0.06
Lung Normal	0.02
Lung Normal	0.01
Lung Normal	0.10
Lung Normal	0.01
Lung Tumor	0.59
Lung Tumor	0.18
Lung Tumor	0.24
Lung Tumor	0.04
Lung Tumor	0.78
Lung Tumor	0.37
Lung Tumor	0.16

[407] Many tested lung tumor samples in Table 1 (6 of 7) expressed 21953 mRNA at higher levels than did normal lung tumor samples. Similarly, a number of breast tumor samples expressed 21953 mRNA to a greater extent than did normal breast tumor samples.

Table 2.

Sample	Relative Expression
Colon Normal	0.00
Colon Normal	0.02
Colon Normal	0.05
Colon Normal	0.01
Colon Tumor	0.03
Colon Tumor	0.24
Colon Tumor	0.07
Colon Tumor	0.03
Colon Tumor	0.03
Colon Tumor	0.04
Liver Metastatic	0.07
Liver Metastatic	0.16
Liver Metastatic	0.23
Liver Normal	0.05
Liver Normal	0.19
Brain Normal	1.50
Brain Normal	0.98
Astrocyte	0.37
Brain Tumor	0.04
Brain Tumor	0.10
Brain Tumor	0.04
Brain Tumor	0.13
HMVEC-Arr	0.22
HMVEC-Prol	0.26
Placenta	0.11
Fetal Adrenal	0.15
Fetal Adrenal	0.18
Fetal Liver	0.71
Fetal Liver	0.18

[408] The mRNA expression data for 21953 mRNA tabulated in Table 2 indicated that (1) 21953 mRNA can be overexpressed in some colon tumor samples relative to normal colon tissue samples; (2) 21953 mRNA is well expressed in metastatic liver samples; (3) 21953 mRNA is highly expressed in normal brain tissue (e.g., increased expression relative to brain tumors), astrocytes, and fetal liver; and (4) 21953 mRNA is also expressed in HMVEC (human microvascular endothelial cells), and fetal adrenal cells.

Table 3.

Sample	Relative Expression
Aorta / normal	0.00
Fetal heart/ normal	2.42
Heart normal	0.66
Heart/ CHF	0.72
Vein/ Normal	0.13
SMC (Aortic)	0.89
Spinal cord/ Normal	0.66
Brain cortex/ Normal	5.94
Brain hypothalamus/ Normal	4.13
Glial cells (Astrocytes)	1.35
Brain/ Glioblastoma	1.12
Breast / Normal	0.18
Breast tumor/ IDC	0.38
Ovary/ Normal	0.39
Ovary/ Tumor	0.16
Pancreas	0.25
Prostate/ Normal	0.18
Prostate/ Tumor	0.15
Colon/ normal	0.07
Colon/tumor	0.56
Colon/IBD	0.10
Kidney/ normal	0.71
Liver/ normal	0.10
Liver fibrosis	0.22
Fetal Liver/normal	2.21
Lung / normal	0.16
Lung/ tumor	0.39
Lung/ COPD	0.22
Spleen/ normal	0.14
Tonsil/ normal	0.11
Lymph node/ normal	0.27
Thymus/ normal	1.16
Epithelial Cells (prostate)	2.04
Endothelial Cells (aortic)	0.27
Skeletal Muscle/ Normal	1.22
Fibroblasts (Dermal)	0.18
Skin/ normal	0.35
Adipose/ Normal	0.06
Osteoblasts (primary)	0.44
Osteoblasts (Undiff)	0.32
Osteoblasts(Diff)	0.29
Osteoclasts	0.08
Aortic SMC Early	1.27
Aortic SMC Late	2.61
shear HUVEC	3.39
static HUVEC	2.14

[409] The mRNA expression data for 21953 mRNA tabulated in Table 3 indicated that 21953 mRNA is highly expressed, for example, in fetal heart, brain cortex, brain hypothalamus, fetal liver, epithelial cells from prostate, aortic smooth muscle cells, and human umbilical vein endothelial cells under both shear and static conditions.

Table 4

Sample	Relative Expression
MCF-7 Breast Tumor	15.15
ZR75 Breast Tumor	6.11
T47D Breast Tumor	1.50
MDA 231 Breast Tumor	0.01
MDA 435 Breast Tumor	0.00
DLD 1 ColonT (stageC)	22.33
SW480 ColonT (stageB)	0.06
SW620 ColonT (stageC)	5.23
HCT116	0.63
HT29	0.01
Colo 205	0.00
NCIH125	0.75
NCIH69	23.28
NCIH322	20.91
NCIH460	1.25
A549	7.11
NHBE	0.83
SKOV-3 ovary	0.22
OVCAR-3 ovary	17.28
293 ovary	44.97
293T ovary	59.75
A549 t6	0.83
A549 t9	1.27
A549 t18	14.63
A549 t24	1.99

[410] Tumor cell lines were xenografted into nude mice. Expression of human 21953 mRNA in tumors harvested from the mice was analyzed using TaqMan. Results are tabulated in Table 4 (excepting the final four rows, see below). The results indicated that, for example, 21953 mRNA is highly expressed in some xenografted colon tumor samples (colonT), some xenografted breast tumor samples, and xenografted ovarian cell lines.

[411] The final four rows of Table 4 tabulate relative 21953 mRNA expression in samples of A549 human lung cancer cells at various hourly time points (time in hours being

indicated with the prefix t) after release from aphidocolin treatment. 21953 mRNA expression peaked at the G1 to S phase transition.

Table 5

Sample	Relative Expression
PIT 337 Colon Normal	0.28
CHT 410 Colon Normal	0.03
CHT 425 Colon Normal	0.13
CHT 371 Colon Normal	0.03
CHT 414 Colonic ACA-B	0.16
CHT 841 Colonic ACA-B	0.07
CHT 807 Colonic ACA-B	0.21
CHT 382 Colonic ACA-B	0.32
CHT 596 Colonic ACA-C	0.00
CHT 907 Colonic ACA-C	0.13
CHT 372 Colonic ACA-C	0.49
NDR 210 Colonic ACA-C	0.13
<u>CHT 1365 Colonic ACA-C</u>	<u>0.03</u>
CLN 741 Liver Normal	0.00
NDR 165 Liver Normal	0.00
NDR 150-Liver Normal	0.06
PIT 236 Liver Normal	0.00
CHT 077-Col-Liver-Metastasis	0.06
CHT-119 Col Liver Metastasis	4.79
CHT 131 Col-Liver Metastasis	0.76
CHT 218 Col Liver Metastasis	1.12
CHT 739 Col Liver Metastasis	0.18
<u>CHT 215 Col Abdominal Metastasis</u>	<u>0.01</u>

[412] 21953 mRNA is cell cycle regulated in the lung carcinoma cell line A549. A549 cells were synchronized with aphidocholin, and then released. mRNA was prepared from the cells at regular intervals after release. 21953 expression peaked during the G1 to S phase transition.

[413] In situ hybridization experiments which provided additional confirmatory results are tabulated in Table 6. 21953 mRNA was observed by in situ hybridization in lung small cell carcinoma and differentiated tumors, but not in normal lung tissue. Similarly, by this analysis, 21953 mRNA expression was elevated in colon tumor samples (2 of 2), metastatic colon tumor samples (2 of 2), and in a differentiated papillary ovarian tumor sample. 21953 mRNA was also detected in normal breast tissue (1 of 1), normal ovarian tissue (1 of 1), and ovarian tumors (2 of 2).

Table 6.

Tissue	Diagnosis	Results
Breast	Normal	+
Breast	Intraductal Carcinoma	-
Colon	Normal	-
Colon	Normal	-
Colon	Tumor	+
Colon	Tumor	+
Colon	Metastasis	+
Colon	Metastasis	++
Liver	Normal	-
Lung	Normal	-
Lung	Small Cell Carcinoma	++
Lung	Differentiated	++
Lung	Differentiated	+/-
Lung	Differentiated	++
Ovary	Normal	+
Ovary	Tumor (well differentiated carcinoma)	+
Ovary	Tumor (moderately differentiated papillary)	++

Example 3: Recombinant Expression of 21953 in Bacterial Cells

[414] In this example, 21953 is expressed as a recombinant glutathione-S-transferase (GST) fusion polypeptide in *E. coli* and the fusion polypeptide is isolated and characterized. Specifically, 21953 is fused to GST and this fusion polypeptide is expressed in *E. coli*, e.g., strain PEB199. Expression of the GST-21953 fusion protein in PEB199 is induced with IPTG. The recombinant fusion polypeptide is purified from crude bacterial lysates of the induced PEB199 strain by affinity chromatography on glutathione beads. Using polyacrylamide gel electrophoretic analysis of the polypeptide purified from the bacterial lysates, the molecular weight of the resultant fusion polypeptide is determined.

Example 4: Expression of Recombinant 21953 Protein in COS Cells

[415] To express the 21953 gene in COS cells, the pcDNA/Amp vector by Invitrogen Corporation (San Diego, CA) is used. This vector contains an SV40 origin of replication, an ampicillin resistance gene, an *E. coli* replication origin, a CMV promoter followed by a polylinker region, and an SV40 intron and polyadenylation site. A DNA fragment encoding the entire 21953 protein and an HA tag (Wilson *et al.* (1984) *Cell* 37:767) or a FLAG tag

fused in-frame to its 3' end of the fragment is cloned into the polylinker region of the vector, thereby placing the expression of the recombinant protein under the control of the CMV promoter.

[416] To construct the plasmid, the 21953 DNA sequence is amplified by PCR using two primers. The 5' primer contains the restriction site of interest followed by approximately twenty nucleotides of the 21953 coding sequence starting from the initiation codon; the 3' end sequence contains complementary sequences to the other restriction site of interest, a translation stop codon, the HA tag or FLAG tag and the last 20 nucleotides of the 21953 coding sequence. The PCR amplified fragment and the pCDNA/Amp vector are digested with the appropriate restriction enzymes and the vector is dephosphorylated using the CIAP enzyme (New England Biolabs, Beverly, MA). Preferably the two restriction sites chosen are different so that the 21953 gene is inserted in the correct orientation. The ligation mixture is transformed into *E. coli* cells (strains HB101, DH5 $\alpha$ , SURE, available from Stratagene Cloning Systems, La Jolla, CA, can be used), the transformed culture is plated on ampicillin media plates, and resistant colonies are selected. Plasmid DNA is isolated from transformants and examined by restriction analysis for the presence of the correct fragment.

[417] COS cells are subsequently transfected with the 21953-pcDNA/Amp plasmid DNA using the calcium phosphate or calcium chloride co-precipitation methods, DEAE-dextran-mediated transfection, lipofection, or electroporation. Other suitable methods for transfecting host cells can be found in Sambrook, J., Fritsh, E. F., and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*. 2nd, ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY. The expression of the 21953 polypeptide is detected by radiolabelling ( $^{35}$ S-methionine or  $^{35}$ S-cysteine available from NEN, Boston, MA, can be used) and immunoprecipitation (Harlow, E. and Lane, D. (1988) *Antibodies: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY) using an HA specific monoclonal antibody. Briefly, the cells are labeled for 8 hours with  $^{35}$ S-methionine (or  $^{35}$ S-cysteine). The culture media are then collected and the cells are lysed using detergents (RIPA buffer, 150 mM NaCl, 1% NP-40, 0.1% SDS, 0.5% DOC, 50 mM Tris, pH 7.5). Both the cell lysate and the culture media are precipitated with an HA specific monoclonal antibody. Precipitated polypeptides are then analyzed by SDS-PAGE.

[418] Alternatively, DNA containing the 21953 coding sequence is cloned directly into the polylinker of the pCDNA/Amp vector using the appropriate restriction sites. The resulting plasmid is transfected into COS cells in the manner described above, and the expression of the 21953 polypeptide is detected by radiolabelling and immunoprecipitation using a 21953 specific monoclonal antibody.

**Equivalents**

[419] Those skilled in the art will recognize, or be able to ascertain using no more than routine experimentation, many equivalents to the specific embodiments of the invention described herein. Such equivalents are intended to be encompassed by the following claims.

1    What is claimed is:

- 1        1. An isolated nucleic acid molecule selected from the group consisting of:
  - 2            a) a nucleic acid molecule comprising a nucleotide sequence which is at least  
3        80% identical to the nucleotide sequence of SEQ ID NO:1, or SEQ ID NO:3;
  - 4            b) a nucleic acid molecule comprising a fragment of at least 800 nucleotides of  
5        the nucleotide sequence of SEQ ID NO:1, or SEQ ID NO:3;
  - 6            c) a nucleic acid molecule which encodes a polypeptide comprising the amino  
7        acid sequence of SEQ ID NO:2;
  - 8            d) a nucleic acid molecule which encodes a fragment of a polypeptide  
9        comprising the amino acid sequence of SEQ ID NO:2, wherein the fragment comprises at  
10      least 15 contiguous amino acids of SEQ ID NO:2; and
  - 11          e) a nucleic acid molecule which encodes a naturally occurring allelic variant of  
12      a polypeptide comprising the amino acid sequence of SEQ ID NO:2, wherein the nucleic  
13      acid molecule hybridizes to a nucleic acid molecule comprising SEQ ID NO:1, 3, or a  
14      complement thereof, under stringent conditions.
- 1        2. The isolated nucleic acid molecule of claim 1, which is selected from the  
2        group consisting of:
  - 3            a) a nucleic acid comprising the nucleotide sequence of SEQ ID NO:1, SEQ ID  
4        NO:3; and
  - 5            b) a nucleic acid molecule which encodes a polypeptide comprising the amino  
6        acid sequence of SEQ ID NO:2.
- 1        3. The nucleic acid molecule of claim 1 further comprising a vector nucleic acid  
2        sequence.
- 1        4. The nucleic acid molecule of claim 1 further comprising a nucleic acid  
2        sequence encoding a heterologous polypeptide.
- 1        5. A host cell which contains the nucleic acid molecule of claim 1.
- 1        6. An isolated polypeptide selected from the group consisting of:

2           a)     a polypeptide which is encoded by a nucleic acid molecule comprising a  
3     nucleotide sequence which is at least 85% identical to a nucleic acid comprising the  
4     nucleotide sequence of SEQ ID NO:1 or SEQ ID NO:3;  
5           b)     a naturally occurring allelic variant of a polypeptide comprising the amino  
6     acid sequence of SEQ ID NO:2, wherein the polypeptide is encoded by a nucleic acid  
7     molecule which hybridizes to a nucleic acid molecule comprising SEQ ID NO:1, SEQ ID  
8     NO:3, or a complement thereof under stringent conditions; and  
9           c)     a fragment of a polypeptide comprising the amino acid sequence of SEQ ID  
10    NO:2, wherein the fragment comprises at least 15 contiguous amino acids of SEQ ID NO:2.

1           7.     The isolated polypeptide of claim 6 comprising the amino acid sequence of  
2     SEQ ID NO:2.

1           8.     The polypeptide of claim 6 further comprising a heterologous amino acid  
2     sequence.

1           9.     An antibody which selectively binds to a polypeptide of claim 6.

1           10.    A method for producing a polypeptide selected from the group consisting of  
2           a)     a polypeptide comprising the amino acid sequence of SEQ ID NO:2;  
3           b)     a polypeptide comprising a fragment of the amino acid sequence of SEQ ID  
4     NO:2, wherein the fragment comprises at least 15 contiguous amino acids of SEQ ID NO:2;  
5     and  
6           c)     a naturally occurring allelic variant of a polypeptide comprising the amino  
7     acid sequence of SEQ ID NO:2, wherein the polypeptide is encoded by a nucleic acid  
8     molecule which hybridizes to a nucleic acid molecule comprising SEQ ID NO:1, SEQ ID  
9     NO:3, or a complement thereof under stringent conditions;  
10        the method, comprising culturing the host cell of claim 5 under conditions in which  
11     the nucleic acid molecule is expressed.

1           11.    A method for detecting the presence of a polypeptide of claim 6 in a sample,  
2     comprising:

3           a)     contacting the sample with a compound which selectively binds to a  
4     polypeptide of claim 6; and  
5           b)     determining whether the compound binds to the polypeptide in the sample.

1           12.    The method of claim 11, wherein the compound is an antibody.

1           13.    A kit comprising a compound which selectively binds to a polypeptide of  
2     claim 6 and instructions for use.

1           14.    A method for detecting the presence of a nucleic acid molecule of claim 1 in  
2     a sample, comprising the steps of:

3           a)     contacting the sample with a nucleic acid probe or primer which selectively  
4     hybridizes to the nucleic acid molecule; and  
5           b)     determining whether the nucleic acid probe or primer binds to a nucleic acid  
6     molecule in the sample.

1           15.    A kit comprising a compound which selectively hybridizes to a nucleic acid  
2     molecule of claim 1 and instructions for use.

1           16.    A method for identifying a compound which binds to a polypeptide of  
2     claim 6 comprising the steps of:  
3           a)     contacting a polypeptide, or a cell expressing a polypeptide of claim 6 with a  
4     test compound; and  
5           b)     determining whether the polypeptide binds to the test compound.

1           17.    The method of claim 16, wherein the binding of the test compound to the  
2     polypeptide is detected by a method selected from the group consisting of:  
3           a)     detection of binding by direct detecting of test compound/polypeptide  
4     binding;  
5           b)     detection of binding using a competition binding assay;  
6           c)     detection of binding using an assay for peptide cleavage.

1           18.    A method for modulating the activity or expression of a polypeptide of claim  
2     6 comprising contacting a polypeptide or a cell expressing a polypeptide of claim 6 with a

3 compound which binds to the polypeptide in a sufficient concentration to modulate the  
4 activity or expression of the polypeptide.

1        19. A method for identifying a compound which modulates the activity of a  
2 polypeptide of claim 6, comprising:  
3            a) contacting a polypeptide of claim 6 with a test compound; and  
4            b) determining the effect of the test compound on the proteolytic activity of the  
5 polypeptide to thereby identify a compound which modulates the proteolytic activity of the  
6 polypeptide.

1        20. A method of treating or preventing a disorder characterized by aberrant  
2 activity or expression of a 21953 nucleic acid or polypeptide in a subject, the method  
3 comprising administering to the subject an effective amount of an agent that modulates the  
4 activity or expression of a 21953 nucleic acid or polypeptide such that the disorder is  
5 ameliorated or prevented.

1        21. A method of modulating the activity of a 21953-expressing cell, comprising  
2 contacting the cell with an amount of an agent that modulates the activity or expression of a  
3 21953 nucleic acid or polypeptide such that the activity of the cell is modulated.

1        22. The method of claim 21, wherein the disorder is a cellular proliferative or  
2 differentiative disorder.

1        23. The method of claim 22, wherein the disorder is lung cancer, colon cancer, or  
2 breast cancer.

1        24. The method of claim 21, wherein the agent is a peptide, a phosphopeptide, a  
2 small molecule, an antibody, or any combination thereof.

1        25. The method of claim 21, wherein the agent is an antisense, a ribozyme, a  
2 triple helix molecule, a 21953 nucleic acid, or any combination thereof.

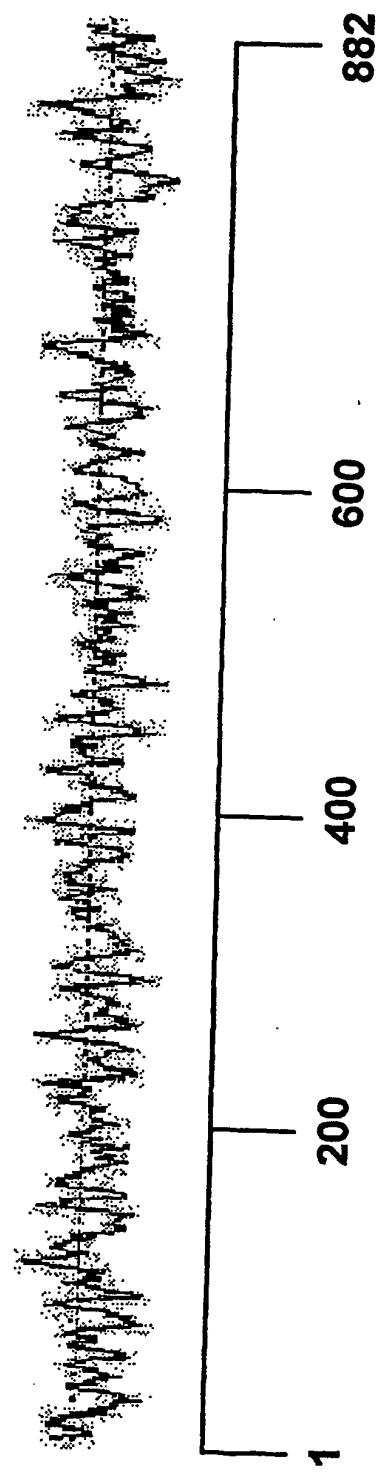
1           26. A method of evaluating a subject for the presence of a cell proliferative  
2 disorder, the method comprising:  
3           a) providing a sample from the subject;  
4           b) detecting an expression level of the nucleic acid of claim 1 in the sample;  
5           c) comparing the expression level to a reference expression level, wherein an  
6 increased expression level relative to the reference expression level is an indication of a cell  
7 proliferative disorder.

1           27. The method of claim 26 wherein the sample comprises cells from lung, breast,  
2 ovarian, prostate, or colonic tissue.

1           28. A method of evaluating a subject for the presence of a cell proliferative disorder,  
2 the method comprising:  
3           a) providing a sample from the subject;  
4           b) detecting an expression level of the polypeptide of claim 6 in the sample;  
5           c) comparing the abundance of the polypeptide to a reference expression level,  
6 wherein an increased abundance relative to the reference expression level is an indication of  
7 a cell proliferative disorder.

1           29. The method of claim 28 wherein the sample comprises cells from lung, breast,  
2 ovarian, prostate, or colonic tissue.

1/4



**FIG. 1**

2/4

**Prolyl Oligopeptidase Domain****from 672 to 744: score 38.1, E = 1.6e-10**

SEQ ID NO:4	*->	vas1linhrGgiyAvvdIRGGeuyGqkwheagtrrxrllknefnDfiaAA
	+++1	+G +++v d RG+ G k+ a + ++ e+D++
21953	672	RINTLASLIGVYVVVIDNNGSCHRLKTEGAFKYKMGQIEIDDQVEGL 718
eylsk1.GytspkriaifGgSngGIL <-*		
	+y1 + + +r+ i	G+S+GG+L
21953	719	QYLASRYDFIDLDRVGIAHGSYGGYL 744

**FIG. 2**

**FIG. 3A**

4/4

DPP IV 21953	(SEQ ID NO:5)	SNEEGYRHICYFQIDKKDCFITKGTTWEVIG-----IEALTSIDLYYISNEYKGMPGGRN 430 RSSGGGLPAPSDFKCPIKEELAITSGEWEVILGRHGSNIQDEVRLVYFEGTKD-SPLEHH 535
		.. * . * : * * . * *** : * * : .. * . * : .. * ::
DPP IV 21953		LYKIQQLSDTYTKVTCLSCELNPERCQYYSVSFSKEAKYYQLRCRSGPGLPLYTLHSSVNDKG 490 LYVSYVNPGEVTRLTDRGYSHSSCCISQHCDFFISKSQNQNP-HCVSLYKLSSPEDDPT 594
		* * : . ; * * : . * . * : * * : . * . * : * . * : *
DPP IV 21953		LRVLEDNSALDKMLQNVQ--MPSKKLDFILLNETKFWYQMLPPFDKSKKYPLLLDVYA 548 CKTKEFWATILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLQPGKKYPTVLFIYG 654
		.. * : : : . * : . * : . : * . : * . : * . : * . : * . : *
DPP IV 21953		GPCSQOKADTVFR--LNWATYLASTENIIVASFDGRGSGYQGDKIMHAIRRLGTFEVEDQ 606 GPQVQLVNNREFGVKYFRINTLASSLGYVVVVDNNGSCHRGKLFEGAFKYYKMGQLEIDQ 714
		* * : . * : . : * . : * . : * . : * . : * . : * . : * . : *
DPP IV 21953		IEAAQQFS-KMGFDVNKRRIAIWGSYGGVTSMVLGSGSGVERCGIAVAPVSRWEYYDSV 665 VEGLQYLASRYDFIDLDRGVIGHGSYGGYLSQLMAMQRSDIFRVALAGAPVTLWIFIYDTG 774
		: * . : . * : . * : . * : * . : * . : * . : * . : * . : * . : *
DPP IV 21953		YTERYMGGLPTPEDNLHDYRNSTVMSRAENFKOVEYILLIHTADDNVHFQOQSQIISKALVD 725 YTERYMGHPDQNEQGYYLGSVAMQAEKFPSEPNRLLLHGELDENVHFAHTSILLSFLVR 834
		* * * * * : * : : . : . : * . : * . : * . : * . : * . : * . : *
DPP IV 21953		VGVDFQAMWYTEDDHGIASSTAHQHIYTHMSHFIKOCEFLP----- 766 AGKPYDLQIYPQERHSIRVPESEGHYELHILHYLQENLGSRIAALKVI 882
		. * : . * : * . * : . : * . : * . : * . : * . : * . : *

**FIG. 3B**



**This Page is Inserted by IFW Indexing and Scanning  
Operations and is not part of the Official Record**

## **BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- BLACK BORDERS**
- IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**
- FADED TEXT OR DRAWING**
- BLURRED OR ILLEGIBLE TEXT OR DRAWING**
- SKEWED/SLANTED IMAGES**
- COLOR OR BLACK AND WHITE PHOTOGRAPHS**
- GRAY SCALE DOCUMENTS**
- LINES OR MARKS ON ORIGINAL DOCUMENT**
- REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**
- OTHER:** \_\_\_\_\_

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.**



(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
27 December 2001 (27.12.2001)

PCT

(10) International Publication Number  
**WO 01/98468 A2**

(51) International Patent Classification<sup>7</sup>: C12N 9/00

(21) International Application Number: PCT/US01/19178

(22) International Filing Date: 13 June 2001 (13.06.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/212,336	16 June 2000 (16.06.2000)	US
60/213,955	22 June 2000 (22.06.2000)	US
60/215,396	29 June 2000 (29.06.2000)	US
60/216,821	7 July 2000 (07.07.2000)	US
60/218,946	14 July 2000 (14.07.2000)	US

(71) Applicant (for all designated States except US): INCYTE GENOMICS, INC. [US/US]; 3160 Porter Drive, Palo Alto, CA 94304 (US).



WO 01/98468 A2

(72) Inventors; and

(73) Inventors/Applicants (for US only): YUE, Henry [US/US]; 826 Lois Avenue, Sunnyvale, CA 94087 (US). ELLIOTT, Vicki, S. [US/US]; 3770 Polton Place Way, San Jose, CA 95121 (US). GANDHI, Ameena, R. [US/US]; 837 Roble Avenue, #1, Menlo Park, CA 94025 (US). LAI, Preeti [IN/US]; P.O. Box 5142, Santa Clara, CA 95056 (US). AU-YOUNG, Janice [US/US]; 233 Golden Eagle Lane, Brisbane, CA 94005 (US). TRIBOULEY, Catherine, M. [FR/US]; 1121 Tennessee Street, #5, San Francisco, CA 94107 (US). DELEGEANE, Angelo, M. [US/US]; 594 Angus Drive, Milpitas, CA 95035 (US). BAUGHN, Mariah, R. [US/US]; 14244 Santiago Road, San Leandro, CA 94577 (US). NGUYEN, Dannie, B. [US/US]; 1403 Ridgewood Drive, San Jose, CA 95118 (US). LEE, Ernestine, A. [US/US]; 624 Kains Street, Albany, CA 94706 (US). HAFALIA, April [US/US]; 2227 Calle de Primavera, Santa Clara, CA 95054 (US). KHAN, Farrah, A. [IN/US]; 3617 Central Road #102, Glenview, IL 60025 (US). WALIA, Narinder, K. [US/US]; 890 Davis Street #205, San Leandro, CA 94577 (US). YAO, Monique, G. [US/US]; 111 Frederick Court, Mountain View, CA 94043 (US). LU, Dyung, Aina, M. [US/US];

233 Coy Drive, San Jose, CA 95123 (US). PATTERSON, Chandra [US/US]; 490 Sherwood Way #1, Menlo Park, CA 94025 (US). TANG, Y., Tom [US/US]; 4230 Ranwick Court, San Jose, CA 95118 (US). WALSH, Roderick, T. [IE/US]; 8 Boundary Court, St. Lawrence Road, Canterbury, Kent CT1 3EZ (GB). AZIMZAI, Yalda [US/US]; 5518 Boulder Canyon Drive, Castro Valley, CA 94552 (US). LU, Yan [CN/US]; 3885 Corrina Way, Palo Alto, CA 94303 (US). RAMKUMAR, Jayalaxmi [IN/US]; 34359 Maybird Circle, Fremont, CA 94555 (US). XU, Yuming [US/US]; 1739 Walnut Drive, Mountain View, CA 94040 (US). REDDY, Roopa [IN/US]; 1233 W. McKinley Avenue #3, Sunnyvale, CA 94086 (US). DAS, Debopriya [IN/US]; 1179 Bonita Avenue, Apt. 3, Mountain View, CA 94040 (US). KEARNEY, Liam [IE/US]; 50 Woodside Avenue, San Francisco, CA 94127 (US). KALLICK, Deborah, A. [US/US]; 900 Olive Street, Menlo Park, CA 94025 (US).

(74) Agents: HAMLET-COX, Diana et al.; Incyte Genomics, Inc., 3160 Porter Drive, Palo Alto, CA 94304 (US).

(81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:

— without international search report and to be republished upon receipt of that report

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: PROTEASES

(57) Abstract: The invention provides human proteases (PRTS) and polynucleotides which identify and encode PRTS. The invention also provides expression vectors, host cells, antibodies, agonists, and antagonists. The invention also provides methods for diagnosing, treating, or preventing disorders associated with aberrant expression of PRTS.

## PROTEASES

### TECHNICAL FIELD

This invention relates to nucleic acid and amino acid sequences of proteases and to the use of  
5 these sequences in hydrolysis of peptide bonds and in the diagnosis, treatment, and prevention of  
gastrointestinal, cardiovascular, autoimmune/inflammatory, cell proliferative, developmental, epithelial,  
neurological, and reproductive disorders, and in the assessment of the effects of exogenous compounds  
on the expression of nucleic acid and amino acid sequences of proteases.

10

### BACKGROUND OF THE INVENTION

Proteases cleave proteins and peptides at the peptide bond that forms the backbone of the  
protein or peptide chain. Proteolysis is one of the most important and frequent enzymatic reactions that  
occurs both within and outside of cells. Proteolysis is responsible for the activation and maturation of  
nascent polypeptides, the degradation of misfolded and damaged proteins, and the controlled turnover of  
15 peptides within the cell. Proteases participate in digestion, endocrine function, and tissue remodeling  
during embryonic development, wound healing, and normal growth. Proteases can play a role in  
regulatory processes by affecting the half life of regulatory proteins. Proteases are involved in the  
etiology or progression of disease states such as inflammation, angiogenesis, tumor dispersion and  
metastasis, cardiovascular disease, neurological disease, and bacterial, parasitic, and viral infections.

20 Proteases can be categorized on the basis of where they cleave their substrates. Exopeptidases,  
which include aminopeptidases, dipeptidyl peptidases, tripeptidases, carboxypeptidases, peptidyl-di-  
peptidases, dipeptidases, and omega peptidases, cleave residues at the termini of their substrates.  
Endopeptidases, including serine proteases, cysteine proteases, and metalloproteases, cleave at residues  
within the peptide. Four principal categories of mammalian proteases have been identified based on  
25 active site structure, mechanism of action, and overall three-dimensional structure. (See Beynon, R.J.  
and J.S. Bond (1994) Proteolytic Enzymes: A Practical Approach, Oxford University Press, New York  
NY, pp. 1-5.)

#### Serine Proteases

The serine proteases (SPs) are a large, widespread family of proteolytic enzymes that include  
30 the digestive enzymes trypsin and chymotrypsin, components of the complement and blood-clotting  
cascades, and enzymes that control the degradation and turnover of macromolecules within the cell and  
in the extracellular matrix. Most of the more than 20 subfamilies can be grouped into six clans, each  
with a common ancestor. These six clans are hypothesized to have descended from at least four  
evolutionarily distinct ancestors. SPs are named for the presence of a serine residue found in the active

catalytic site of most families. The active site is defined by the catalytic triad, a set of conserved asparagine, histidine, and serine residues critical for catalysis. These residues form a charge relay network that facilitates substrate binding. Other residues outside the active site form an oxyanion hole that stabilizes the tetrahedral transition intermediate formed during catalysis. SPs have a wide range of substrates and can be subdivided into subfamilies on the basis of their substrate specificity. The main subfamilies are named for the residue(s) after which they cleave: trypases (after arginine or lysine), aspases (after aspartate), chymases (after phenylalanine or leucine), metases (methionine), and serases (after serine) (Rawlings, N.D. and A.J. Barrett (1994) Methods Enzymol. 244:19-61).

Most mammalian serine proteases are synthesized as zymogens, inactive precursors that are activated by proteolysis. For example, trypsinogen is converted to its active form, trypsin, by enteropeptidase. Enteropeptidase is an intestinal protease that removes an N-terminal fragment from trypsinogen. The remaining active fragment is trypsin, which in turn activates the precursors of the other pancreatic enzymes. Likewise, proteolysis of prothrombin, the precursor of thrombin, generates three separate polypeptide fragments. The N-terminal fragment is released while the other two fragments, which comprise active thrombin, remain associated through disulfide bonds.

The two largest SP subfamilies are the chymotrypsin (S1) and subtilisin (S8) families. Some members of the chymotrypsin family contain two structural domains unique to this family. Kringle domains are triple-looped, disulfide cross-linked domains found in varying copy number. Kringles are thought to play a role in binding mediators such as membranes, other proteins or phospholipids, and in the regulation of proteolytic activity (PROSITE PDOC00020). Apple domains are 90 amino-acid repeated domains, each containing six conserved cysteines. Three disulfide bonds link the first and sixth, second and fifth, and third and fourth cysteines (PROSITE PDOC00376). Apple domains are involved in protein-protein interactions. S1 family members include trypsin, chymotrypsin, coagulation factors IX-XII, complement factors B, C, and D, granzymes, kallikrein, and tissue- and urokinase-plasminogen activators. The subtilisin family has members found in the eubacteria, archaebacteria, eukaryotes, and viruses. Subtilisins include the proprotein-processing endopeptidases kexin and furin and the pituitary prohormone convertases PC1, PC2, PC3, PC6, and PACE4 (Rawlings and Barrett, *supra*). The prolyl oligopeptidase (S9) family includes enzymes from prokaryotes and eukaryotes with greatly differing specificities. Dipeptidyl peptidase IV (DPP-IV) is identical to CD26 and is implicated in the inactivation of peptide hormones, as well as in regulating T-cell growth (reviewed in Kahne, T. et al. (1999) Int. J. Mol. Med. 4:3-15; Mentlein, R. (1999) Regul. Pept. 85:9-24). Inhibition of DPP-IV has been suggested as a treatment for type 2 diabetes (Holst, J.J. and C.F. Deacon (1998) Diabetes 47:1663-1670), and lowered serum DPP-IV activity has been measured in anorexia and bulimia patients (van West, D. et al. (2000) Eur. Arch. Psych. Clin. Neurosci. 250:86-92).

SPs have functions in many normal processes and some have been implicated in the etiology or treatment of disease. Enterokinase, the initiator of intestinal digestion, is found in the intestinal brush border, where it cleaves the acidic propeptide from trypsinogen to yield active trypsin (Kitamoto, Y. et al. (1994) Proc. Natl. Acad. Sci. USA 91:7588-7592). Prolylcarboxypeptidase, a lysosomal serine peptidase that cleaves peptides such as angiotensin II and III and [des-Arg9] bradykinin, shares sequence homology with members of both the serine carboxypeptidase and prolylendopeptidase families (Tan, F. et al. (1993) J. Biol. Chem. 268:16631-16638). The protease neutropsin may influence synapse formation and neuronal connectivity in the hippocampus in response to neural signaling (Chen, Z.-L. et al. (1995) J. Neurosci. 15:5088-5097). Tissue plasminogen activator is useful for acute management of stroke (Zivin, J.A. (1999) Neurology 53:14-19) and myocardial infarction (Ross, A.M. (1999) Clin. Cardiol. 22:165-171). Some receptors (PAR, for proteinase-activated receptor), highly expressed throughout the digestive tract, are activated by proteolytic cleavage of an extracellular domain. The major agonists for PARs, thrombin, trypsin, and mast cell tryptase, are released in allergy and inflammatory conditions. Control of PAR activation by proteases has been suggested as a promising therapeutic target (Vergnolle, N. (2000) Aliment. Pharmacol. Ther. 14:257-266; Rice, K.D. et al. (1998) Curr. Pharm. Des. 4:381-396). Prostate-specific antigen (PSA) is a kallikrein-like serine protease synthesized and secreted exclusively by epithelial cells in the prostate gland. Serum PSA is elevated in prostate cancer and is the most sensitive physiological marker for monitoring cancer progression and response to therapy. PSA can also identify the prostate as the origin of a metastatic tumor (Brawer, M.K. and P.H. Lange (1989) Urology 33:11-16).

The signal peptidase is a specialized class of SP found in all prokaryotic and eukaryotic cell types that serves in the processing of signal peptides from certain proteins. Signal peptides are amino-terminal domains of a protein which direct the protein from its ribosomal assembly site to a particular cellular or extracellular location. Once the protein has been exported, removal of the signal sequence by a signal peptidase and posttranslational processing, e.g., glycosylation or phosphorylation, activate the protein. Signal peptidases exist as multi-subunit complexes in both yeast and mammals. The canine signal peptidase complex is composed of five subunits, all associated with the microsomal membrane and containing hydrophobic regions that span the membrane one or more times (Shelness, G.S. and G. Blobel (1990) J. Biol. Chem. 265:9512-9519). Some of these subunits serve to fix the complex in its proper position on the membrane while others contain the actual catalytic activity.

Another family of proteases which have a serine in their active site are dependent on the hydrolysis of ATP for their activity. These proteases contain proteolytic core domains and regulatory ATPase domains which can be identified by the presence of the P-loop, an ATP/GTP-binding motif (PROSITE PDOC00803). Members of this family include the eukaryotic mitochondrial matrix

proteases, Clp protease and the proteasome. Clp protease was originally found in plant chloroplasts but is believed to be widespread in both prokaryotic and eukaryotic cells. The gene for early-onset torsion dystonia encodes a protein related to Clp protease (Ozelius, L.J. et al. (1998) *Adv. Neurol.* 78:93-105).

The proteasome is an intracellular protease complex found in some bacteria and in all 5 eukaryotic cells, and plays an important role in cellular physiology. Proteasomes are associated with the ubiquitin conjugation system (UCS), a major pathway for the degradation of cellular proteins of all types, including proteins that function to activate or repress cellular processes such as transcription and cell cycle progression (Ciechanover, A. (1994) *Cell* 79:13-21). In the UCS pathway, proteins targeted for degradation are conjugated to ubiquitin, a small heat stable protein. The ubiquitinated protein is 10 then recognized and degraded by the proteasome. The resultant ubiquitin-peptide complex is hydrolyzed by a ubiquitin carboxyl terminal hydrolase, and free ubiquitin is released for reutilization by the UCS. Ubiquitin-proteasome systems are implicated in the degradation of mitotic cyclic kinases, oncoproteins, tumor suppressor genes (p53), cell surface receptors associated with signal transduction, transcriptional regulators, and mutated or damaged proteins (Ciechanover, *supra*). This pathway has 15 been implicated in a number of diseases, including cystic fibrosis, Angelman's syndrome, and Liddle syndrome (reviewed in Schwartz, A.L. and A. Ciechanover (1999) *Annu. Rev. Med.* 50:57-74). A murine proto-oncogene, *Ubp*, encodes a nuclear ubiquitin protease whose overexpression leads to oncogenic transformation of NIH3T3 cells. The human homologue of this gene is consistently elevated in small cell tumors and adenocarcinomas of the lung (Gray, D.A. (1995) *Oncogene* 10:2179-2183). 20 Ubiquitin carboxyl-terminal-hydrolase is involved in the differentiation of a lymphoblastic leukemia cell line to a non-dividing mature state (Maki, A. et al. (1996) *Differentiation* 60:59-66). In neurons, ubiquitin carboxyl terminal hydrolase (PGP 9.5) expression is strong in the abnormal structures that occur in human neurodegenerative diseases (Lowe, J. et al. (1990) *J. Pathol.* 161:153-160). The proteasome is a large (~2000 kDa) multisubunit complex composed of a central catalytic core 25 containing a variety of proteases arranged in four seven-membered rings with the active sites facing inwards into the central cavity, and terminal ATPase subunits covering the outer port of the cavity and regulating substrate entry (for review, see Schmidt, M. et al. (1999) *Curr. Opin. Chem. Biol.* 3:584-591).

#### Cysteine Proteases

30 Cysteine proteases (CPs) are involved in diverse cellular processes ranging from the processing of precursor proteins to intracellular degradation. Nearly half of the CPs known are present only in viruses. CPs have a cysteine as the major catalytic residue at the active site where catalysis proceeds via a thioester intermediate and is facilitated by nearby histidine and asparagine residues. A glutamine residue is also important, as it helps to form an oxyanion hole. Two important CP families include the

papain-like enzymes (C1) and the calpains (C2). Papain-like family members are generally lysosomal or secreted and therefore are synthesized with signal peptides as well as propeptides. Most members bear a conserved motif in the propeptide that may have structural significance (Karrer, K.M. et al. (1993) Proc. Natl. Acad. Sci. USA 90:3063-3067). Three-dimensional structures of papain family

5 members show a bilobed molecule with the catalytic site located between the two lobes. Papains include cathepsins B, C, H, L, and S, certain plant allergens and dipeptidyl peptidase (for a review, see Rawlings, N.D. and A.J. Barrett (1994) Methods Enzymol. 244:461-486).

Some CPs are expressed ubiquitously, while others are produced only by cells of the immune system. Of particular note, CPs are produced by monocytes, macrophages and other cells which 10 migrate to sites of inflammation and secrete molecules involved in tissue repair. Overabundance of these repair molecules plays a role in certain disorders. In autoimmune diseases such as rheumatoid arthritis, secretion of the cysteine peptidase cathepsin C degrades collagen, laminin, elastin and other structural proteins found in the extracellular matrix of bones. Bone weakened by such degradation is also more susceptible to tumor invasion and metastasis. Cathepsin L expression may also contribute to 15 the influx of mononuclear cells which exacerbates the destruction of the rheumatoid synovium (Keyszer, G.M. (1995) Arthritis Rheum. 38:976-984).

Calpains are calcium-dependent cytosolic endopeptidases which contain both an N-terminal catalytic domain and a C-terminal calcium-binding domain. Calpain is expressed as a proenzyme heterodimer consisting of a catalytic subunit unique to each isoform and a regulatory subunit common 20 to different isoforms. Each subunit bears a calcium-binding EF-hand domain. The regulatory subunit also contains a hydrophobic glycine-rich domain that allows the enzyme to associate with cell membranes. Calpains are activated by increased intracellular calcium concentration, which induces a change in conformation and limited autolysis. The resultant active molecule requires a lower calcium concentration for its activity (Chan, S.L. and M.P. Mattson (1999) J. Neurosci. Res. 58:167-190). 25 Calpain expression is predominantly neuronal, although it is present in other tissues. Several chronic neurodegenerative disorders, including ALS, Parkinson's disease and Alzheimer's disease are associated with increased calpain expression (Chan and Mattson, supra). Calpain-mediated breakdown of the cytoskeleton has been proposed to contribute to brain damage resulting from head injury (McCracken, E. et al. (1999) J. Neurotrauma 16:749-761). Calpain-3 is predominantly expressed in 30 skeletal muscle, and is responsible for limb-girdle muscular dystrophy type 2A (Minami, N. et al. (1999) J. Neurol. Sci. 171:31-37).

Another family of thiol proteases is the caspases, which are involved in the initiation and execution phases of apoptosis. A pro-apoptotic signal can activate initiator caspases that trigger a proteolytic caspase cascade, leading to the hydrolysis of target proteins and the classic apoptotic death

of the cell. Two active site residues, a cysteine and a histidine, have been implicated in the catalytic mechanism. Caspases are among the most specific endopeptidases, cleaving after aspartate residues. Caspases are synthesized as inactive zymogens consisting of one large (p20) and one small (p10) subunit separated by a small spacer region, and a variable N-terminal prodomain. This prodomain 5 interacts with cofactors that can positively or negatively affect apoptosis. An activating signal causes autoproteolytic cleavage of a specific aspartate residue (D297 in the caspase-1 numbering convention) and removal of the spacer and prodomain, leaving a p10/p20 heterodimer. Two of these heterodimers interact via their small subunits to form the catalytically active tetramer. The long prodomains of some caspase family members have been shown to promote dimerization and auto-processing of procaspases. 10 Some caspases contain a "death effector domain" in their prodomain by which they can be recruited into self-activating complexes with other caspases and FADD protein associated death receptors or the TNF receptor complex. In addition, two dimers from different caspase family members can associate, changing the substrate specificity of the resultant tetramer. Endogenous caspase inhibitors (inhibitor of apoptosis proteins, or IAPs) also exist. All these interactions have clear effects on the control of 15 apoptosis (reviewed in Chan and Mattson, *supra*; Salveson, G.S. and V.M. Dixit (1999) Proc. Natl. Acad. Sci. USA 96:10964-10967).

Caspases have been implicated in a number of diseases. Mice lacking some caspases have severe nervous system defects due to failed apoptosis in the neuroepithelium and suffer early lethality. Others show severe defects in the inflammatory response, as caspases are responsible for processing IL-1<sup>b</sup> and possibly other inflammatory cytokines (Chan and Mattson, *supra*). Cowpox virus and baculoviruses target caspases to avoid the death of their host cell and promote successful infection. In addition, increases in inappropriate apoptosis have been reported in AIDS, neurodegenerative diseases and ischemic injury; while a decrease in cell death is associated with cancer (Salveson and Dixit, *supra*; Thompson, C.B. (1995) Science 267:1456-1462).

## 25 Aspartyl proteases

Aspartyl proteases (APs) include the lysosomal proteases cathepsins D and E, as well as chymosin, renin, and the gastric pepsins. Most retroviruses encode an AP, usually as part of the pol polyprotein. APs, also called acid proteases, are monomeric enzymes consisting of two domains, each domain containing one half of the active site with its own catalytic aspartic acid residue. APs are most 30 active in the range of pH 2-3, at which one of the aspartate residues is ionized and the other neutral. The pepsin family of APs contains many secreted enzymes, and all are likely to be synthesized with signal peptides and propeptides. Most family members have three disulfide loops, the first ~5 residue loop following the first aspartate, the second 5-6 residue loop preceding the second aspartate, and the third and largest loop occurring toward the C terminus. Retropepsins, on the other hand, are analogous

to a single domain of pepsin, and become active as homodimers with each retropepsin monomer contributing one half of the active site. Retropepsins are required for processing the viral polyproteins.

APs have roles in various tissues, and some have been associated with disease. Renin mediates the first step in processing the hormone angiotensin, which is responsible for regulating electrolyte balance and blood pressure (reviewed in Crews, D.E. and S.R. Williams (1999) *Hum. Biol.* 71:475-503). Abnormal regulation and expression of cathepsins are evident in various inflammatory disease states. Expression of cathepsin D is elevated in synovial tissues from patients with rheumatoid arthritis and osteoarthritis. The increased expression and differential regulation of the cathepsins are linked to the metastatic potential of a variety of cancers (Chambers, A.F. et al. (1993) *Crit. Rev. Oncol.* 10: 4:95-114).

#### Metalloproteases

Metalloproteases require a metal ion for activity, usually manganese or zinc. Examples of manganese metalloenzymes include aminopeptidase P and human proline dipeptidase (PEPD). Aminopeptidase P can degrade bradykinin, a nonapeptide activated in a variety of inflammatory responses. Aminopeptidase P has been implicated in coronary ischemia/reperfusion injury. Administration of aminopeptidase P inhibitors has been shown to have a cardioprotective effect in rats (Ersahin, C. et al (1999) *J. Cardiovasc. Pharmacol.* 34:604-611).

Most zinc-dependent metalloproteases share a common sequence in the zinc-binding domain. The active site is made up of two histidines which act as zinc ligands and a catalytic glutamic acid C-terminal to the first histidine. Proteins containing this signature sequence are known as the metzincins and include aminopeptidases B and N, angiotensin-converting enzyme, neurolysin, the matrix metalloproteases and the adamalysins (ADAMS). An alternate sequence is found in the zinc carboxypeptidases, in which all three conserved residues – two histidines and a glutamic acid – are involved in zinc binding.

A number of the neutral metalloendopeptidases, including angiotensin converting enzyme and the aminopeptidases, are involved in the metabolism of peptide hormones. High aminopeptidase B activity, for example, is found in the adrenal glands and neurohypophyses of hypertensive rats (Prieto, I. et al. (1998) *Horm. Metab. Res.* 30:246-248). Oligopeptidase M/neurolysin can hydrolyze bradykinin as well as neuropeptides (Serizawa, A. et al. (1995) *J. Biol. Chem.* 270:2092-2098).

Neurotensin is a vasoactive peptide that can act as a neurotransmitter in the brain, where it has been implicated in limiting food intake (Tritos, N.A. et al. (1999) *Neuropeptides* 33:339-349).

The matrix metalloproteases (MMPs) are a family of at least 23 enzymes that can degrade components of the extracellular matrix (ECM). They are Zn<sup>+2</sup> endopeptidases with an N-terminal catalytic domain. Nearly all members of the family have a hinge peptide and C-terminal domain which

can bind to substrate molecules in the ECM or to inhibitors produced by the tissue (TIMPs, for tissue inhibitor of metalloprotease; Campbell, I.L. et al. (1999) Trends Neurosci. 22:285). The presence of fibronectin-like repeats, transmembrane domains, or C-terminal hemopexinase-like domains can be used to separate MMPs into collagenase, gelatinase, stromelysin and membrane-type MMP subfamilies. In 5 the inactive form, the Zn<sup>+2</sup> ion in the active site interacts with a cysteine in the pro-sequence. Activating factors disrupt the Zn<sup>+2</sup>-cysteine interaction, or "cysteine switch," exposing the active site. This partially activates the enzyme, which then cleaves off its propeptide and becomes fully active. MMPs are often activated by the serine proteases plasmin and furin. MMPs are often regulated by stoichiometric, noncovalent interactions with inhibitors; the balance of protease to inhibitor, then, is 10 very important in tissue homeostasis (reviewed in Yong, V.W. et al. (1998) Trends Neurosci. 21:75). Ehlers-Danlos syndrome type VII C is caused by mutations in the procollagen I N-proteinase gene (Colige, A. et al. (1999) Am. J. Hum. Genet. 65:308-317).

MMPs are implicated in a number of diseases including osteoarthritis (Mitchell, P. et al. (1996) J. Clin. Invest. 97:761), atherosclerotic plaque rupture (Sukhova, G.K. et al. (1999) Circulation 15 99:2503), aortic aneurysm (Schneiderman, J. et al. (1998) Am. J. Path. 152:703), non-healing wounds (Saarialho-Kere, U.K. et al. (1994) J. Clin. Invest. 94:79), bone resorption (Blavier, L. and J.M. Delaisse (1995) J. Cell. Sci. 108:3649), age-related macular degeneration (Steen, B. et al. (1998) Invest. Ophthalmol. Vis. Sci. 39:2194), emphysema (Finlay, G.A. et al. (1997) Thorax 52:502), myocardial infarction (Rohde, L.E. et al. (1999) Circulation 99:3063) and dilated cardiomyopathy (Thomas, C.V. et al. (1998) Circulation 97:1708). MMP inhibitors prevent metastasis of mammary carcinoma and experimental tumors in rat, and Lewis lung carcinoma, hemangioma, and human ovarian carcinoma xenografts in mice (Eccles, S.A. et al. (1996) Cancer Res. 56:2815; Anderson et al. (1996) Cancer Res. 56:715-718; Volpert, O.V. et al. (1996) J. Clin. Invest. 98:671; Taraboletti, G. et al. (1995) J. NCI 87:293; Davies, B. et al. (1993) Cancer Res. 53:2087). MMPs may be active in Alzheimer's disease. 25 A number of MMPs are implicated in multiple sclerosis, and administration of MMP inhibitors can relieve some of its symptoms (reviewed in Yong, supra).

Another family of metalloproteases is the ADAMs, for A Disintegrin and Metalloprotease Domain, which they share with their close relatives the adamalysins, snake venom metalloproteases (SVMPs). ADAMs combine features of both cell surface adhesion molecules and proteases, containing 30 a prodomain, a protease domain, a disintegrin domain, a cysteine rich domain, an epidermal growth factor repeat, a transmembrane domain, and a cytoplasmic tail. The first three domains listed above are also found in the SVMPs. The ADAMs possess four potential functions: proteolysis, adhesion, signaling and fusion. The ADAMs share the metzincin zinc binding sequence and are inhibited by some MMP antagonists such as TIMP-1.

ADAMs are implicated in such processes as sperm-egg binding and fusion, myoblast fusion, and protein-ectodomain processing or shedding of cytokines, cytokine receptors, adhesion proteins and other extracellular protein domains (Schlöndorff, J. and C.P. Blobel (1999) *J. Cell. Sci.* 112:3603-3617). The Kuzbanian protein cleaves a substrate in the NOTCH pathway (possibly NOTCH itself), 5 activating the program for lateral inhibition in *Drosophila* neural development. Two ADAMs, TACE (ADAM 17) and ADAM 10, are proposed to have analogous roles in the processing of amyloid precursor protein in the brain (Schlöndorff and Blobel, *supra*). TACE has also been identified as the TNF activating enzyme (Black, R.A. et al. (1997) *Nature* 385:729). TNF is a pleiotropic cytokine that is important in mobilizing host defenses in response to infection or trauma, but can cause severe 10 damage in excess and is often overproduced in autoimmune disease. TACE cleaves membrane-bound pro-TNF to release a soluble form. Other ADAMs may be involved in a similar type of processing of other membrane-bound molecules.

The ADAMTS sub-family has all of the features of ADAM family metalloproteases and contain an additional thrombospondin domain (TS). The prototypic ADAMTS was identified in mouse, 15 found to be expressed in heart and kidney and upregulated by proinflammatory stimuli (Kuno, K. et al. (1997) *J. Biol. Chem.* 272:556-562). To date eleven members are recognized by the Human Genome Organization (HUGO; <http://www.gene.ucl.ac.uk/users/hester/adamts.html#Approved>). Members of this family have the ability to degrade aggrecan, a high molecular weight proteoglycan which provides cartilage with important mechanical properties including compressibility, and which is lost during the 20 development of arthritis. Enzymes which degrade aggrecan are thus considered attractive targets to prevent and slow the degradation of articular cartilage (See, e.g., Tortorella, M.D. (1999) *Science* 284:1664; Abbaszade, I. (1999) *J. Biol. Chem.* 274:23443). Other members are reported to have antiangiogenic potential (Kuno et al., *supra*) and/or procollagen processing (Colige, A. et al. (1997) *Proc. Natl. Acad. Sci. USA* 94:2374).

25 The discovery of new proteases and the polynucleotides encoding them satisfies a need in the art by providing new compositions which are useful in hydrolysis of peptide bonds and in the diagnosis, prevention, and treatment of gastrointestinal, cardiovascular, autoimmune/inflammatory, cell proliferative, developmental, epithelial, neurological, and reproductive disorders, and in the assessment of the effects of exogenous compounds on the expression of nucleic acid and amino acid sequences of 30 proteases.

## SUMMARY OF THE INVENTION

The invention features purified polypeptides, proteases, referred to collectively as "PRTS" and individually as "PRTS-1," "PRTS-2," "PRTS-3," "PRTS-4," "PRTS-5," "PRTS-6," "PRTS-7,"

"PRTS-8," "PRTS-9," "PRTS-10," "PRTS-11," "PRTS-12," "PRTS-13," "PRTS-14," "PRTS-15," "PRTS-16," "PRTS-17," "PRTS-18," "PRTS-19," "PRTS-20," and "PRTS-21." In one aspect, the invention provides an isolated polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, b) a 5 polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-21. In one alternative, the invention provides an isolated polypeptide 10 comprising the amino acid sequence of SEQ ID NO:1-21.

The invention further provides an isolated polynucleotide encoding a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-15 21; c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-21; and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-21. In one alternative, the polynucleotide encodes a polypeptide selected from the group consisting of SEQ ID NO:1-21. In another alternative, the polynucleotide is selected from the group consisting of SEQ ID NO:22-42.

20 Additionally, the invention provides a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, c) a biologically 25 active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-21. In one alternative, the invention provides a cell transformed with the recombinant polynucleotide. In another alternative, the invention provides a transgenic organism comprising the recombinant polynucleotide.

30 The invention also provides a method for producing a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group

consisting of SEQ ID NO:1-21, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-21. The method comprises a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide comprising a promoter sequence operably linked to a 5 polynucleotide encoding the polypeptide, and b) recovering the polypeptide so expressed.

Additionally, the invention provides an isolated antibody which specifically binds to a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group 10 consisting of SEQ ID NO:1-21, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-21.

The invention further provides an isolated polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID 15 NO:22-42, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:22-42, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). In one alternative, the polynucleotide comprises at least 60 contiguous nucleotides.

20 Additionally, the invention provides a method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:22-42, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:22-42, c) a 25 polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). The method comprises a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said 30 probe and said target polynucleotide or fragments thereof, and b) detecting the presence or absence of said hybridization complex, and optionally, if present, the amount thereof. In one alternative, the probe comprises at least 60 contiguous nucleotides.

The invention further provides a method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide selected from the group consisting of a) a

polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:22-42, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:22-42, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the 5 polynucleotide of b), and e) an RNA equivalent of a)-d). The method comprises a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof.

The invention further provides a composition comprising an effective amount of a polypeptide 10 selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, and d) an immunogenic fragment of a polypeptide 15 having an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, and a pharmaceutically acceptable excipient. In one embodiment, the composition comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-21. The invention additionally provides a method of treating a disease or condition associated with decreased expression of functional PRTS, comprising administering to a patient in need of such treatment the composition.

20 The invention also provides a method for screening a compound for effectiveness as an agonist of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, c) a biologically active fragment of a polypeptide having an 25 amino acid sequence selected from the group consisting of SEQ ID NO:1-21, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-21. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting agonist activity in the sample. In one alternative, the invention provides a composition comprising an agonist compound identified by the method and a pharmaceutically acceptable excipient. In another alternative, the invention provides a method of treating a disease or condition associated with decreased expression of functional PRTS, comprising administering to a 30 patient in need of such treatment the composition.

Additionally, the invention provides a method for screening a compound for effectiveness as an antagonist of a polypeptide selected from the group consisting of a) a polypeptide comprising an

amino acid sequence selected from the group consisting of SEQ ID NO:1-21, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, and

5 d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-21. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting antagonist activity in the sample. In one alternative, the invention provides a composition comprising an antagonist compound identified by the method and a pharmaceutically acceptable excipient. In another alternative, the invention provides a method of

10 treating a disease or condition associated with overexpression of functional PRTS, comprising administering to a patient in need of such treatment the composition.

The invention further provides a method of screening for a compound that specifically binds to a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-21. The method comprises a) combining the polypeptide with at least one test compound under suitable conditions, and b) detecting binding of the polypeptide to the test compound, thereby identifying a compound that specifically binds to the polypeptide.

The invention further provides a method of screening for a compound that modulates the activity of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-21. The method comprises a) combining the polypeptide with at least one test compound under conditions permissive for the activity of the polypeptide, b) assessing the activity of the polypeptide in the presence of the test compound, and c) comparing the activity of the polypeptide in the presence of the test compound with the activity of the polypeptide in the absence of the test compound, wherein a change in the activity of the polypeptide in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide.

The invention further provides a method for screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a sequence selected from the group consisting of SEQ ID NO:22-42, the method comprising a) exposing a sample comprising the target polynucleotide to a compound, and b) detecting altered expression of the target polynucleotide.

The invention further provides a method for assessing toxicity of a test compound, said method comprising a) treating a biological sample containing nucleic acids with the test compound; b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide selected from the group consisting of i) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:22-42, ii) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:22-42, iii) a polynucleotide having a sequence complementary to i), iv) a polynucleotide complementary to the polynucleotide of ii), and v) an RNA equivalent of i)-iv). Hybridization occurs under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide selected from the group consisting of i) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:22-42, ii) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:22-42, iii) a polynucleotide complementary to the polynucleotide of i), iv) a polynucleotide complementary to the polynucleotide of ii); and v) an RNA equivalent of i)-iv). Alternatively, the target polynucleotide comprises a fragment of a polynucleotide sequence selected from the group consisting of i)-v) above; c) quantifying the amount of hybridization complex; and d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

#### BRIEF DESCRIPTION OF THE TABLES

Table 1 summarizes the nomenclature for the full length polynucleotide and polypeptide sequences of the present invention.

Table 2 shows the GenBank identification number and annotation of the nearest GenBank homolog for polypeptides of the invention. The probability score for the match between each polypeptide and its GenBank homolog is also shown.

Table 3 shows structural features of polypeptide sequences of the invention, including predicted motifs and domains, along with the methods, algorithms, and searchable databases used for analysis of the polypeptides.

Table 4 lists the cDNA and/or genomic DNA fragments which were used to assemble 5 polynucleotide sequences of the invention, along with selected fragments of the polynucleotide sequences.

Table 5 shows the representative cDNA library for polynucleotides of the invention.

Table 6 provides an appendix which describes the tissues and vectors used for construction of the cDNA libraries shown in Table 5.

10 Table 7 shows the tools, programs, and algorithms used to analyze the polynucleotides and polypeptides of the invention, along with applicable descriptions, references, and threshold parameters.

#### DESCRIPTION OF THE INVENTION

Before the present proteins, nucleotide sequences, and methods are described, it is understood 15 that this invention is not limited to the particular machines, materials and methods described, as these may vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to limit the scope of the present invention which will be limited only by the appended claims.

It must be noted that as used herein and in the appended claims, the singular forms "a," "an," 20 and "the" include plural reference unless the context clearly dictates otherwise. Thus, for example, a reference to "a host cell" includes a plurality of such host cells, and a reference to "an antibody" is a reference to one or more antibodies and equivalents thereof known to those skilled in the art, and so forth.

Unless defined otherwise, all technical and scientific terms used herein have the same meanings 25 as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any machines, materials, and methods similar or equivalent to those described herein can be used to practice or test the present invention, the preferred machines, materials and methods are now described. All publications mentioned herein are cited for the purpose of describing and disclosing the cell lines, protocols, reagents and vectors which are reported in the publications and which might be used in 30 connection with the invention. Nothing herein is to be construed as an admission that the invention is not entitled to antedate such disclosure by virtue of prior invention.

**DEFINITIONS**

“PRTS” refers to the amino acid sequences of substantially purified PRTS obtained from any species, particularly a mammalian species, including bovine, ovine, porcine, murine, equine, and 5 human, and from any source, whether natural, synthetic, semi-synthetic, or recombinant.

The term “agonist” refers to a molecule which intensifies or mimics the biological activity of PRTS. Agonists may include proteins, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of PRTS either by directly interacting with PRTS or by acting on components of the biological pathway in which PRTS participates.

10 An “allelic variant” is an alternative form of the gene encoding PRTS. Allelic variants may result from at least one mutation in the nucleic acid sequence and may result in altered mRNAs or in polypeptides whose structure or function may or may not be altered. A gene may have none, one, or many allelic variants of its naturally occurring form. Common mutational changes which give rise to allelic variants are generally ascribed to natural deletions, additions, or substitutions of nucleotides.  
15 Each of these types of changes may occur alone, or in combination with the others, one or more times in a given sequence.

“Altered” nucleic acid sequences encoding PRTS include those sequences with deletions, insertions, or substitutions of different nucleotides, resulting in a polypeptide the same as PRTS or a polypeptide with at least one functional characteristic of PRTS. Included within this definition are 20 polymorphisms which may or may not be readily detectable using a particular oligonucleotide probe of the polynucleotide encoding PRTS, and improper or unexpected hybridization to allelic variants, with a locus other than the normal chromosomal locus for the polynucleotide sequence encoding PRTS. The encoded protein may also be “altered,” and may contain deletions, insertions, or substitutions of amino acid residues which produce a silent change and result in a functionally equivalent PRTS. Deliberate 25 amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues, as long as the biological or immunological activity of PRTS is retained. For example, negatively charged amino acids may include aspartic acid and glutamic acid, and positively charged amino acids may include lysine and arginine. Amino acids with uncharged polar side chains having similar hydrophilicity values may 30 include: asparagine and glutamine; and serine and threonine. Amino acids with uncharged side chains having similar hydrophilicity values may include: leucine, isoleucine, and valine; glycine and alanine; and phenylalanine and tyrosine.

The terms “amino acid” and “amino acid sequence” refer to an oligopeptide, peptide, polypeptide, or protein sequence, or a fragment of any of these, and to naturally occurring or synthetic

molecules. Where "amino acid sequence" is recited to refer to a sequence of a naturally occurring protein molecule, "amino acid sequence" and like terms are not meant to limit the amino acid sequence to the complete native amino acid sequence associated with the recited protein molecule.

"Amplification" relates to the production of additional copies of a nucleic acid sequence.

5 Amplification is generally carried out using polymerase chain reaction (PCR) technologies well known in the art.

The term "antagonist" refers to a molecule which inhibits or attenuates the biological activity of PRTS. Antagonists may include proteins such as antibodies, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of PRTS either by  
10 directly interacting with PRTS or by acting on components of the biological pathway in which PRTS participates.

The term "antibody" refers to intact immunoglobulin molecules as well as to fragments thereof, such as Fab, F(ab')<sub>2</sub>, and Fv fragments, which are capable of binding an epitopic determinant.

Antibodies that bind PRTS polypeptides can be prepared using intact polypeptides or using fragments  
15 containing small peptides of interest as the immunizing antigen. The polypeptide or oligopeptide used to immunize an animal (e.g., a mouse, a rat, or a rabbit) can be derived from the translation of RNA, or synthesized chemically, and can be conjugated to a carrier protein if desired. Commonly used carriers that are chemically coupled to peptides include bovine serum albumin, thyroglobulin, and keyhole limpet hemocyanin (KLH). The coupled peptide is then used to immunize the animal.

20 The term "antigenic determinant" refers to that region of a molecule (i.e., an epitope) that makes contact with a particular antibody. When a protein or a fragment of a protein is used to immunize a host animal, numerous regions of the protein may induce the production of antibodies which bind specifically to antigenic determinants (particular regions or three-dimensional structures on the protein). An antigenic determinant may compete with the intact antigen (i.e., the immunogen used to  
25 elicit the immune response) for binding to an antibody.

The term "antisense" refers to any composition capable of base-pairing with the "sense" (coding) strand of a specific nucleic acid sequence. Antisense compositions may include DNA; RNA; peptide nucleic acid (PNA); oligonucleotides having modified backbone linkages such as phosphorothioates, methylphosphonates, or benzylphosphonates; oligonucleotides having modified

30 sugar groups such as 2'-methoxyethyl sugars or 2'-methoxyethoxy sugars; or oligonucleotides having modified bases such as 5-methyl cytosine, 2'-deoxyuracil, or 7-deaza-2'-deoxyguanosine. Antisense molecules may be produced by any method including chemical synthesis or transcription. Once introduced into a cell, the complementary antisense molecule base-pairs with a naturally occurring nucleic acid sequence produced by the cell to form duplexes which block either transcription or

translation. The designation "negative" or "minus" can refer to the antisense strand, and the designation "positive" or "plus" can refer to the sense strand of a reference DNA molecule.

The term "biologically active" refers to a protein having structural, regulatory, or biochemical functions of a naturally occurring molecule. Likewise, "immunologically active" or "immunogenic" 5 refers to the capability of the natural, recombinant, or synthetic PRTS, or of any oligopeptide thereof, to induce a specific immune response in appropriate animals or cells and to bind with specific antibodies.

"Complementary" describes the relationship between two single-stranded nucleic acid sequences that anneal by base-pairing. For example, 5'-AGT-3' pairs with its complement, 10 3'-TCA-5'.

A "composition comprising a given polynucleotide sequence" and a "composition comprising a given amino acid sequence" refer broadly to any composition containing the given polynucleotide or amino acid sequence. The composition may comprise a dry formulation or an aqueous solution. Compositions comprising polynucleotide sequences encoding PRTS or fragments of PRTS may be 15 employed as hybridization probes. The probes may be stored in freeze-dried form and may be associated with a stabilizing agent such as a carbohydrate. In hybridizations, the probe may be deployed in an aqueous solution containing salts (e.g., NaCl), detergents (e.g., sodium dodecyl sulfate; SDS), and other components (e.g., Denhardt's solution, dry milk, salmon sperm DNA, etc.).

"Consensus sequence" refers to a nucleic acid sequence which has been subjected to repeated 20 DNA sequence analysis to resolve uncalled bases, extended using the XL-PCR kit (Applied Biosystems, Foster City CA) in the 5'-and/or the 3' direction, and resequenced, or which has been assembled from one or more overlapping cDNA, EST, or genomic DNA fragments using a computer program for fragment assembly, such as the GELVIEW fragment assembly system (GCG, Madison WI) or Phrap (University of Washington, Seattle WA). Some sequences have been both extended and assembled to 25 produce the consensus sequence.

"Conservative amino acid substitutions" are those substitutions that are predicted to least interfere with the properties of the original protein, i.e., the structure and especially the function of the protein is conserved and not significantly changed by such substitutions. The table below shows amino acids which may be substituted for an original amino acid in a protein and which are regarded as 30 conservative amino acid substitutions.

Original Residue	Conservative Substitution
Ala	Gly, Ser
Arg	His, Lys
Asn	Asp, Gln, His
Asp	Asn, Glu

	Cys	Ala, Ser
	Gln	Asn, Glu, His
	Glu	Asp, Gln, His
	Gly	Ala
5	His	Asn, Arg, Gln, Glu
	Ile	Leu, Val
	Leu	Ile, Val
	Lys	Arg, Gin, Glu
	Met	Leu, Ile
10	Phe	His, Met, Leu, Trp, Tyr
	Ser	Cys, Thr
	Thr	Ser, Val
	Trp	Phe, Tyr
	Tyr	His, Phe, Trp
15	Val	Ile, Leu, Thr

Conservative amino acid substitutions generally maintain (a) the structure of the polypeptide backbone in the area of the substitution, for example, as a beta sheet or alpha helical conformation, (b) the charge or hydrophobicity of the molecule at the site of the substitution, and/or (c) the bulk of the side chain.

A "deletion" refers to a change in the amino acid or nucleotide sequence that results in the absence of one or more amino acid residues or nucleotides.

The term "derivative" refers to a chemically modified polynucleotide or polypeptide. Chemical modifications of a polynucleotide can include, for example, replacement of hydrogen by an alkyl, acyl, hydroxyl, or amino group. A derivative polynucleotide encodes a polypeptide which retains at least one biological or immunological function of the natural molecule. A derivative polypeptide is one modified by glycosylation, pegylation, or any similar process that retains at least one biological or immunological function of the polypeptide from which it was derived.

A "detectable label" refers to a reporter molecule or enzyme that is capable of generating a measurable signal and is covalently or noncovalently joined to a polynucleotide or polypeptide.

"Differential expression" refers to increased or upregulated; or decreased, downregulated, or absent gene or protein expression, determined by comparing at least two different samples. Such comparisons may be carried out between, for example, a treated and an untreated sample, or a diseased and a normal sample.

35 A "fragment" is a unique portion of PRTS or the polynucleotide encoding PRTS which is identical in sequence to but shorter in length than the parent sequence. A fragment may comprise up to the entire length of the defined sequence, minus one nucleotide/amino acid residue. For example, a fragment may comprise from 5 to 1000 contiguous nucleotides or amino acid residues. A fragment used as a probe, primer, antigen, therapeutic molecule, or for other purposes, may be at least 5, 10,

15, 16, 20, 25, 30, 40, 50, 60, 75, 100, 150, 250 or at least 500 contiguous nucleotides or amino acid residues in length. Fragments may be preferentially selected from certain regions of a molecule. For example, a polypeptide fragment may comprise a certain length of contiguous amino acids selected from the first 250 or 500 amino acids (or first 25% or 50%) of a polypeptide as shown in a certain 5 defined sequence. Clearly these lengths are exemplary, and any length that is supported by the specification, including the Sequence Listing, tables, and figures, may be encompassed by the present embodiments.

A fragment of SEQ ID NO:22-42 comprises a region of unique polynucleotide sequence that specifically identifies SEQ ID NO:22-42, for example, as distinct from any other sequence in the 10 genome from which the fragment was obtained. A fragment of SEQ ID NO:22-42 is useful, for example, in hybridization and amplification technologies and in analogous methods that distinguish SEQ ID NO:22-42 from related polynucleotide sequences. The precise length of a fragment of SEQ ID NO:22-42 and the region of SEQ ID NO:22-42 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

15 A fragment of SEQ ID NO:1-21 is encoded by a fragment of SEQ ID NO:22-42. A fragment of SEQ ID NO:1-21 comprises a region of unique amino acid sequence that specifically identifies ~~SEQ ID NO:1-21~~. For example, a fragment of SEQ ID NO:1-21 is useful as an immunogenic peptide for the development of antibodies that specifically recognize SEQ ID NO:1-21. The precise length of ~~a fragment of SEQ ID NO:1-21 and the region of SEQ ID NO:1-21 to which the fragment~~ 20 ~~corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.~~

A “full-length” polynucleotide sequence is one containing at least a translation initiation codon (e.g., methionine) followed by an open reading frame and a translation termination codon. A “full length” polynucleotide sequence encodes a “full length” polypeptide sequence.

25 “Homology” refers to sequence similarity or, interchangeably, sequence identity, between two or more polynucleotide sequences or two or more polypeptide sequences.

The terms “percent identity” and “% identity,” as applied to polynucleotide sequences, refer to the percentage of residue matches between at least two polynucleotide sequences aligned using a standardized algorithm. Such an algorithm may insert, in a standardized and reproducible way, gaps in 30 the sequences being compared in order to optimize alignment between two sequences, and therefore achieve a more meaningful comparison of the two sequences.

Percent identity between polynucleotide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program. This program is part of the LASERGENE software package, a suite of molecular

biological analysis programs (DNASTAR, Madison WI). CLUSTAL V is described in Higgins, D.G. and P.M. Sharp (1989) CABIOS 5:151-153 and in Higgins, D.G. et al. (1992) CABIOS 8:189-191.

For pairwise alignments of polynucleotide sequences, the default parameters are set as follows:

Ktuple=2, gap penalty=5, window=4, and "diagonals saved"=4. The "weighted" residue weight table is selected as the default. Percent identity is reported by CLUSTAL V as the "percent similarity" between aligned polynucleotide sequences.

Alternatively, a suite of commonly used and freely available sequence comparison algorithms is provided by the National Center for Biotechnology Information (NCBI) Basic Local Alignment Search Tool (BLAST) (Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410), which is available from several sources, including the NCBI, Bethesda, MD, and on the Internet at <http://www.ncbi.nlm.nih.gov/BLAST/>. The BLAST software suite includes various sequence analysis programs including "blastn," that is used to align a known polynucleotide sequence with other polynucleotide sequences from a variety of databases. Also available is a tool called "BLAST 2 Sequences" that is used for direct pairwise comparison of two nucleotide sequences. "BLAST 2 Sequences" can be accessed and used interactively at <http://www.ncbi.nlm.nih.gov/gorf/bl2.html>. The "BLAST 2 Sequences" tool can be used for both blastn and blastp (discussed below). BLAST programs are commonly used with gap and other parameters set to default settings. For example, to compare two nucleotide sequences, one may use blastn with the "BLAST 2 Sequences" tool Version 2.0.12 (April-21-2000) set at default parameters. Such default parameters may be, for example:

20           *Matrix: BLOSUM62*

*Reward for match: 1*

*Penalty for mismatch: -2*

*Open Gap: 5 and Extension Gap: 2 penalties*

*Gap x drop-off: 50*

25           *Expect: 10*

*Word Size: 11*

*Filter: on*

Percent identity may be measured over the length of an entire defined sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over 30 the length of a fragment taken from a larger, defined sequence, for instance, a fragment of at least 20, at least 30, at least 40, at least 50, at least 70, at least 100, or at least 200 contiguous nucleotides. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures, or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

Nucleic acid sequences that do not show a high degree of identity may nevertheless encode similar amino acid sequences due to the degeneracy of the genetic code. It is understood that changes in a nucleic acid sequence can be made using this degeneracy to produce multiple nucleic acid sequences that all encode substantially the same protein.

5       The phrases "percent identity" and "% identity," as applied to polypeptide sequences, refer to the percentage of residue matches between at least two polypeptide sequences aligned using a standardized algorithm. Methods of polypeptide sequence alignment are well-known. Some alignment methods take into account conservative amino acid substitutions. Such conservative substitutions, explained in more detail above, generally preserve the charge and hydrophobicity at the site of  
10      substitution, thus preserving the structure (and therefore function) of the polypeptide.

Percent identity between polypeptide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program (described and referenced above). For pairwise alignments of polypeptide sequences using CLUSTAL V, the default parameters are set as follows: Ktuple=1, gap penalty=3, window=5, and  
15      "diagonals saved"=5. The PAM250 matrix is selected as the default residue weight table. As with polynucleotide alignments, the percent identity is reported by CLUSTAL V as the "percent similarity" between aligned polypeptide sequence pairs.

Alternatively the NCBI BLAST software suite may be used. For example, for a pairwise comparison of two polypeptide sequences, one may use the "BLAST 2 Sequences" tool Version 2.0.12  
20      (April 21, 2000) with blastp set at default parameters. Such default parameters may be, for example:

Matrix: BLOSUM62  
Open Gap: 11 and Extension Gap: 1 penalties  
Gap x drop-off: 50  
Expect: 10  
25      Word Size: 3  
Filter: on  
Percent identity may be measured over the length of an entire defined polypeptide sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined polypeptide sequence, for instance,  
30      a fragment of at least 15, at least 20, at least 30, at least 40, at least 50, at least 70 or at least 150 contiguous residues. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

"Human artificial chromosomes" (HACs) are linear microchromosomes which may contain DNA sequences of about 6 kb to 10 Mb in size and which contain all of the elements required for chromosome replication, segregation and maintenance.

The term "humanized antibody" refers to an antibody molecule in which the amino acid sequence in the non-antigen binding regions has been altered so that the antibody more closely resembles a human antibody, and still retains its original binding ability.

"Hybridization" refers to the process by which a polynucleotide strand anneals with a complementary strand through base pairing under defined hybridization conditions. Specific hybridization is an indication that two nucleic acid sequences share a high degree of complementarity.

10 Specific hybridization complexes form under permissive annealing conditions and remain hybridized after the "washing" step(s). The washing step(s) is particularly important in determining the stringency of the hybridization process, with more stringent conditions allowing less non-specific binding, i.e., binding between pairs of nucleic acid strands that are not perfectly matched. Permissive conditions for annealing of nucleic acid sequences are routinely determinable by one of ordinary skill in the art and 15 may be consistent among hybridization experiments, whereas wash conditions may be varied among experiments to achieve the desired stringency, and therefore hybridization specificity. Permissive annealing conditions occur, for example, at 68°C in the presence of about 6 x SSC, about 1% (w/v) SDS, and about 100 µg/ml sheared, denatured salmon sperm DNA.

Generally, stringency of hybridization is expressed, in part, with reference to the temperature 20 under which the wash step is carried out. Such wash temperatures are typically selected to be about 5°C to 20°C lower than the thermal melting point ( $T_m$ ) for the specific sequence at a defined ionic strength and pH. The  $T_m$  is the temperature (under defined ionic strength and pH) at which 50% of the target sequence hybridizes to a perfectly matched probe. An equation for calculating  $T_m$  and conditions for nucleic acid hybridization are well known and can be found in Sambrook, J. et al. (1989) Molecular Cloning: A Laboratory Manual, 2<sup>nd</sup> ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; specifically see volume 2, chapter 9.

High stringency conditions for hybridization between polynucleotides of the present invention include wash conditions of 68°C in the presence of about 0.2 x SSC and about 0.1% SDS, for 1 hour. Alternatively, temperatures of about 65°C, 60°C, 55°C, or 42°C may be used. SSC concentration may 30 be varied from about 0.1 to 2 x SSC, with SDS being present at about 0.1%. Typically, blocking reagents are used to block non-specific hybridization. Such blocking reagents include, for instance, sheared and denatured salmon sperm DNA at about 100-200 µg/ml. Organic solvent, such as formamide at a concentration of about 35-50% v/v, may also be used under particular circumstances, such as for RNA:DNA hybridizations. Useful variations on these wash conditions will be readily

apparent to those of ordinary skill in the art. Hybridization, particularly under high stringency conditions, may be suggestive of evolutionary similarity between the nucleotides. Such similarity is strongly indicative of a similar role for the nucleotides and their encoded polypeptides.

The term "hybridization complex" refers to a complex formed between two nucleic acid sequences by virtue of the formation of hydrogen bonds between complementary bases. A hybridization complex may be formed in solution (e.g., C<sub>0</sub>t or R<sub>0</sub>t analysis) or formed between one nucleic acid sequence present in solution and another nucleic acid sequence immobilized on a solid support (e.g., paper, membranes, filters, chips, pins or glass slides, or any other appropriate substrate to which cells or their nucleic acids have been fixed).

10 The words "insertion" and "addition" refer to changes in an amino acid or nucleotide sequence resulting in the addition of one or more amino acid residues or nucleotides, respectively.

"Immune response" can refer to conditions associated with inflammation, trauma, immune disorders, or infectious or genetic disease, etc. These conditions can be characterized by expression of various factors, e.g., cytokines, chemokines, and other signaling molecules, which may affect cellular 15 and systemic defense systems.

An "immunogenic fragment" is a polypeptide or oligopeptide fragment of PRTS which is capable of eliciting an immune response when introduced into a living organism, for example, a mammal. The term "immunogenic fragment" also includes any polypeptide or oligopeptide fragment of PRTS which is useful in any of the antibody production methods disclosed herein or known in the art.

20 The term "microarray" refers to an arrangement of a plurality of polynucleotides, polypeptides, or other chemical compounds on a substrate.

The terms "element" and "array element" refer to a polynucleotide, polypeptide, or other chemical compound having a unique and defined position on a microarray.

25 The term "modulate" refers to a change in the activity of PRTS. For example, modulation may cause an increase or a decrease in protein activity, binding characteristics, or any other biological, functional, or immunological properties of PRTS.

The phrases "nucleic acid" and "nucleic acid sequence" refer to a nucleotide, oligonucleotide, polynucleotide, or any fragment thereof. These phrases also refer to DNA or RNA of genomic or synthetic origin which may be single-stranded or double-stranded and may represent the sense or the 30 antisense strand, to peptide nucleic acid (PNA), or to any DNA-like or RNA-like material.

"Operably linked" refers to the situation in which a first nucleic acid sequence is placed in a functional relationship with a second nucleic acid sequence. For instance, a promoter is operably linked to a coding sequence if the promoter affects the transcription or expression of the coding

sequence. Operably linked DNA sequences may be in close proximity or contiguous and, where necessary to join two protein coding regions, in the same reading frame.

"Peptide nucleic acid" (PNA) refers to an antisense molecule or anti-gene agent which comprises an oligonucleotide of at least about 5 nucleotides in length linked to a peptide backbone of 5 amino acid residues ending in lysine. The terminal lysine confers solubility to the composition. PNAs preferentially bind complementary single stranded DNA or RNA and stop transcript elongation, and may be pegylated to extend their lifespan in the cell.

"Post-translational modification" of an PRTS may involve lipidation, glycosylation, phosphorylation, acetylation, racemization, proteolytic cleavage, and other modifications known in the 10 art. These processes may occur synthetically or biochemically. Biochemical modifications will vary by cell type depending on the enzymatic milieu of PRTS.

"Probe" refers to nucleic acid sequences encoding PRTS, their complements, or fragments thereof, which are used to detect identical, allelic or related nucleic acid sequences. Probes are 15 isolated oligonucleotides or polynucleotides attached to a detectable label or reporter molecule. Typical labels include radioactive isotopes, ligands, chemiluminescent agents, and enzymes. "Primers" are short nucleic acids, usually DNA oligonucleotides, which may be annealed to a target polynucleotide by complementary base-pairing. The primer may then be extended along the target DNA strand by a DNA polymerase enzyme. Primer pairs can be used for amplification (and identification) of a nucleic acid sequence, e.g., by the polymerase chain reaction (PCR).

20 Probes and primers as used in the present invention typically comprise at least 15 contiguous nucleotides of a known sequence. In order to enhance specificity, longer probes and primers may also be employed, such as probes and primers that comprise at least 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, or at least 150 consecutive nucleotides of the disclosed nucleic acid sequences. Probes and primers may be considerably longer than these examples, and it is understood that any length supported by the 25 specification, including the tables, figures, and Sequence Listing, may be used.

Methods for preparing and using probes and primers are described in the references, for example Sambrook, J. et al. (1989) Molecular Cloning: A Laboratory Manual, 2<sup>nd</sup> ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; Ausubel, F.M. et al. (1987) Current Protocols in Molecular Biology, Greene Publ. Assoc. & Wiley-Intersciences, New York NY; Innis, M. et al. (1990) PCR 30 Protocols, A Guide to Methods and Applications, Academic Press, San Diego CA. PCR primer pairs can be derived from a known sequence, for example, by using computer programs intended for that purpose such as Primer (Version 0.5, 1991, Whitehead Institute for Biomedical Research, Cambridge MA).

Oligonucleotides for use as primers are selected using software known in the art for such purpose. For example, OLIGO 4.06 software is useful for the selection of PCR primer pairs of up to 100 nucleotides each, and for the analysis of oligonucleotides and larger polynucleotides of up to 5,000 nucleotides from an input polynucleotide sequence of up to 32 kilobases. Similar primer selection programs have incorporated additional features for expanded capabilities. For example, the PrimOU primer selection program (available to the public from the Genome Center at University of Texas South West Medical Center, Dallas TX) is capable of choosing specific primers from megabase sequences and is thus useful for designing primers on a genome-wide scope. The Primer3 primer selection program (available to the public from the Whitehead Institute/MIT Center for Genome Research, Cambridge MA) allows the user to input a "mispriming library," in which sequences to avoid as primer binding sites are user-specified. Primer3 is useful, in particular, for the selection of oligonucleotides for microarrays. (The source code for the latter two primer selection programs may also be obtained from their respective sources and modified to meet the user's specific needs.) The PrimeGen program (available to the public from the UK Human Genome Mapping Project Resource Centre, Cambridge UK) designs primers based on multiple sequence alignments, thereby allowing selection of primers that hybridize to either the most conserved or least conserved regions of aligned nucleic acid sequences. Hence, this program is useful for identification of both unique and conserved oligonucleotides and polynucleotide fragments. The oligonucleotides and polynucleotide fragments identified by any of the above selection methods are useful in hybridization technologies, for example, as PCR or sequencing primers, microarray elements, or specific probes to identify fully or partially complementary polynucleotides in a sample of nucleic acids. Methods of oligonucleotide selection are not limited to those described above.

A "recombinant nucleic acid" is a sequence that is not naturally occurring or has a sequence that is made by an artificial combination of two or more otherwise separated segments of sequence. This artificial combination is often accomplished by chemical synthesis or, more commonly, by the artificial manipulation of isolated segments of nucleic acids, e.g., by genetic engineering techniques such as those described in Sambrook, *supra*. The term recombinant includes nucleic acids that have been altered solely by addition, substitution, or deletion of a portion of the nucleic acid. Frequently, a recombinant nucleic acid may include a nucleic acid sequence operably linked to a promoter sequence. Such a recombinant nucleic acid may be part of a vector that is used, for example, to transform a cell.

Alternatively, such recombinant nucleic acids may be part of a viral vector, e.g., based on a vaccinia virus, that could be used to vaccinate a mammal wherein the recombinant nucleic acid is expressed, inducing a protective immunological response in the mammal.

A "regulatory element" refers to a nucleic acid sequence usually derived from untranslated regions of a gene and includes enhancers, promoters, introns, and 5' and 3' untranslated regions (UTRs). Regulatory elements interact with host or viral proteins which control transcription, translation, or RNA stability.

5        "Reporter molecules" are chemical or biochemical moieties used for labeling a nucleic acid, amino acid, or antibody. Reporter molecules include radionuclides; enzymes; fluorescent, chemiluminescent, or chromogenic agents; substrates; cofactors; inhibitors; magnetic particles; and other moieties known in the art.

An "RNA equivalent," in reference to a DNA sequence, is composed of the same linear  
10 sequence of nucleotides as the reference DNA sequence with the exception that all occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The term "sample" is used in its broadest sense. A sample suspected of containing PRTS, nucleic acids encoding PRTS, or fragments thereof may comprise a bodily fluid; an extract from a cell,  
15 chromosome, organelle, or membrane isolated from a cell; a cell; genomic DNA, RNA, or cDNA, in solution or bound to a substrate; a tissue; a tissue print; etc.

The terms "specific binding" and "specifically binding" refer to that interaction between a protein or peptide and an agonist, an antibody, an antagonist, a small molecule, or any natural or synthetic binding composition. The interaction is dependent upon the presence of a particular structure  
20 of the protein, e.g., the antigenic determinant or epitope, recognized by the binding molecule. For example, if an antibody is specific for epitope "A," the presence of a polypeptide comprising the epitope A, or the presence of free unlabeled A, in a reaction containing free labeled A and the antibody will reduce the amount of labeled A that binds to the antibody.

The term "substantially purified" refers to nucleic acid or amino acid sequences that are  
25 removed from their natural environment and are isolated or separated, and are at least 60% free, preferably at least 75% free, and most preferably at least 90% free from other components with which they are naturally associated.

A "substitution" refers to the replacement of one or more amino acid residues or nucleotides by different amino acid residues or nucleotides, respectively.  
30        "Substrate" refers to any suitable rigid or semi-rigid support including membranes, filters, chips, slides, wafers, fibers, magnetic or nonmagnetic beads, gels, tubing, plates, polymers, microparticles and capillaries. The substrate can have a variety of surface forms, such as wells, trenches, pins, channels and pores, to which polynucleotides or polypeptides are bound.

A "transcript image" refers to the collective pattern of gene expression by a particular cell type or tissue under given conditions at a given time.

"Transformation" describes a process by which exogenous DNA is introduced into a recipient cell. Transformation may occur under natural or artificial conditions according to various methods well known in the art, and may rely on any known method for the insertion of foreign nucleic acid sequences into a prokaryotic or eukaryotic host cell. The method for transformation is selected based on the type of host cell being transformed and may include, but is not limited to, bacteriophage or viral infection, electroporation, heat shock, lipofection, and particle bombardment. The term "transformed cells" includes stably transformed cells in which the inserted DNA is capable of replication either as an autonomously replicating plasmid or as part of the host chromosome, as well as transiently transformed cells which express the inserted DNA or RNA for limited periods of time.

A "transgenic organism," as used herein, is any organism, including but not limited to animals and plants, in which one or more of the cells of the organism contains heterologous nucleic acid introduced by way of human intervention, such as by transgenic techniques well known in the art. The nucleic acid is introduced into the cell, directly or indirectly by introduction into a precursor of the cell, by way of deliberate genetic manipulation, such as by microinjection or by infection with a recombinant virus. The term "genetic manipulation" does not include classical cross-breeding, or in vitro-fertilization, but rather is directed to the introduction of a recombinant DNA molecule. The transgenic organisms contemplated in accordance with the present invention include bacteria, cyanobacteria, fungi, plants and animals. The isolated DNA of the present invention can be introduced into the host by methods known in the art, for example infection, transfection, transformation or transconjugation. Techniques for transferring the DNA of the present invention into such organisms are widely known and provided in references such as Sambrook et al. (1989), supra.

A "variant" of a particular nucleic acid sequence is defined as a nucleic acid sequence having at least 40% sequence identity to the particular nucleic acid sequence over a certain length of one of the nucleic acid sequences using blastn with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of nucleic acids may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% or greater sequence identity over a certain defined length. A variant may be described as, for example, an "allelic" (as defined above), "splice," "species," or "polymorphic" variant. A splice variant may have significant identity to a reference molecule, but will generally have a greater or lesser number of polynucleotides due to alternative splicing of exons during mRNA processing. The corresponding polypeptide may

possess additional functional domains or lack domains that are present in the reference molecule. Species variants are polynucleotide sequences that vary from one species to another. The resulting polypeptides will generally have significant amino acid identity relative to each other. A polymorphic variant is a variation in the polynucleotide sequence of a particular gene between individuals of a given 5 species. Polymorphic variants also may encompass "single nucleotide polymorphisms" (SNPs) in which the polynucleotide sequence varies by one nucleotide base. The presence of SNPs may be indicative of, for example, a certain population, a disease state, or a propensity for a disease state.

A "variant" of a particular polypeptide sequence is defined as a polypeptide sequence having at least 40% sequence identity to the particular polypeptide sequence over a certain length of one of the 10 polypeptide sequences using blastp with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of polypeptides may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% or greater sequence identity over a certain defined length of one of the polypeptides.

15

## THE INVENTION

The invention is based on the discovery of new human proteases (PRTS), the polynucleotides encoding PRTS, and the use of these compositions for the diagnosis, treatment, or prevention of gastrointestinal, cardiovascular, autoimmune/inflammatory, cell proliferative, developmental, epithelial, 20 neurological, and reproductive disorders.

Table 1 summarizes the nomenclature for the full length polynucleotide and polypeptide sequences of the invention. Each polynucleotide and its corresponding polypeptide are correlated to a single Incyte project identification number (Incyte Project ID). Each polypeptide sequence is denoted by both a polypeptide sequence identification number (Polypeptide SEQ ID NO:) and an Incyte 25 polypeptide sequence number (Incyte Polypeptide ID) as shown. Each polynucleotide sequence is denoted by both a polynucleotide sequence identification number (Polynucleotide SEQ ID NO:) and an Incyte Polynucleotide consensus sequence number (Incyte Polynucleotide ID) as shown.

Table 2 shows sequences with homology to the polypeptides of the invention as identified by BLAST analysis against the GenBank protein (genpept) database. Columns 1 and 2 show the 30 polypeptide sequence identification number (Polypeptide SEQ ID NO:) and the corresponding Incyte polypeptide sequence number (Incyte Polypeptide ID) for polypeptides of the invention. Column 3 shows the GenBank identification number (Genbank ID NO:) of the nearest GenBank homolog. Column 4 shows the probability score for the match between each polypeptide and its GenBank

homolog. Column 5 shows the annotation of the GenBank homolog along with relevant citations where applicable, all of which are expressly incorporated by reference herein.

Table 3 shows various structural features of the polypeptides of the invention. Columns 1 and 2 show the polypeptide sequence identification number (SEQ ID NO:) and the corresponding Incyte 5 polypeptide sequence number (Incyte Polypeptide ID) for each polypeptide of the invention. Column 3 shows the number of amino acid residues in each polypeptide. Column 4 shows potential phosphorylation sites, and column 5 shows potential glycosylation sites, as determined by the MOTIFS program of the GCG sequence analysis software package (Genetics Computer Group, Madison WI). Column 6 shows amino acid residues comprising signature sequences, domains, and motifs. Column 7 10 shows analytical methods for protein structure/function analysis and in some cases, searchable databases to which the analytical methods were applied.

Together, Tables 2 and 3 summarize the properties of polypeptides of the invention, and these properties establish that the claimed polypeptides are proteases. For example, SEQ ID NO:1 is a ubiquitin carboxyl terminal hydrolase. SEQ ID NO:1 is 48% identical, from residue M1 to residue 15 G225, to human ubiquitin-specific processing protease (GenBank ID g9971757) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 1.00e-49, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:1 contains a ubiquitin carboxyl-terminal hydrolase catalytic site domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM 20 database of conserved protein family domains. The score is 53.4 bits and the E-value is 4.9e-12, which indicates the probability of obtaining the observed structural motif by chance. The presence of this motif was corroborated by BLIMPS (probability score=2.6e-4) and MOTIFS analyses. This provides further evidence that SEQ ID NO:1 is a ubiquitin carboxyl-terminal hydrolase. In an alternative 25 example, SEQ ID NO:2 is 45% identical to amino acids 15-235 of human prostasin, a serine protease (GenBank ID g1143194) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 1.3e-46, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:2 also contains a trypsin family serine protease active site domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. This 30 match has a probability score of 2.7e-58. BLIMPS, MOTIFS, and PROFILESCAN analyses confirm the presence of this domain. (See Table 3.) BLIMPS analysis also reveals a kringle domain, providing further corroborative evidence that SEQ ID NO:2 is a serine protease of the trypsin family. In an alternative example, SEQ ID NO:7 is a dipeptidase which hydrolyses a variety of peptides (Kozak, E. and S. Tate (1982) J. Biol. Chem. 257:6322-6327), and is responsible for the hydrolysis of the beta

lactam rings of antibiotics such as penem and carbapenem (Campbell et al., (1984) J. Biol. Chem. 259:14586-14590). SEQ ID NO:7 shows 48% amino acid sequence identity over 377 amino acids (total length equals 411 amino acids) to human dipeptidase precursor (GenBank ID g219600) as determined by Basic Local Alignment Search Tool (BLAST). The BLAST probability score is 1.1e-5, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. Additionally, the protease of the invention demonstrates a renal dipeptidase domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. The HMM score for the renal dipeptidase PFAM hit is 412.7. Data from BLIMPS, MOTIFS, BLAST-DOMO, and BLAST-PRODOM analyses provide 10 further corroborative evidence that SEQ ID NO:7 is a renal dipeptidase. The BLIMPS-BLOCKS hit scores for localized regions range from 1040-1537. The BLAST-DOMO hit probability score is 5.2e-85. The BLAST-PRODOM hit probability score is 4.7e-73. In an alternative example, SEQ ID NO:8 is 86% identical to human transmembrane tryptase (GenBank ID g6103629) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 3.9e-166, 15 which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:8 contains a trypsin family protease active site domain with a probability score of 5.3e-89 as determined by searching for matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. BLIMPS, MOTIFS, and PROFILESCAN analyses confirm the presence of this motif. BLIMPS analysis also shows that SEQ ID NO:8 contains a kringle domain and 20 a type I fibronectin domain. HMMER-based analysis reveals the presence of a transmembrane domain (See Table 3.). Taken together, these analyses show that SEQ ID NO:8 is a transmembrane member of the trypsin family of serine proteases. In an alternative example, SEQ ID NO:17 shares 44% local identity with human membrane-type serine protease 1 (MT-SP1, GenBank ID g6002714) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 25 5.1e-94, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:17 contains a trypsin family serine protease active site domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) HMM-based analysis also reveals a transmembrane domain near the N-terminus of SEQ ID NO:17. A domain found in the low-density 30 lipoprotein receptor and other proteins, including MT-SP1 (PDOC00929) was also identified in this way. The presence of the trypsin active site motif is confirmed by PROFILESCAN, BLIMPS, and MOTIFS analyses. BLIMPS analysis revealed the presence of kringle and type I fibronectin domains. Taken together, these data provide further corroborative evidence that SEQ ID NO:17 is a transmembrane member of the trypsin family of serine proteases. SEQ ID NO:3-6, SEQ ID NO:9-16,

and SEQ ID NO:18-21 were analyzed and annotated in a similar manner. The algorithms and parameters for the analysis of SEQ ID NO:1-21 are described in Table 7.

As shown in Table 4, the full length polynucleotide sequences of the present invention were assembled using cDNA sequences or coding (exon) sequences derived from genomic DNA, or any combination of these two types of sequences. Columns 1 and 2 list the polynucleotide sequence identification number (Polynucleotide SEQ ID NO:) and the corresponding Incyte polynucleotide consensus sequence number (Incyte Polynucleotide ID) for each polynucleotide of the invention. Column 3 shows the length of each polynucleotide sequence in basepairs. Column 4 lists fragments of the polynucleotide sequences which are useful, for example, in hybridization or amplification technologies that identify SEQ ID NO:22-42 or that distinguish between SEQ ID NO:22-42 and related polynucleotide sequences. Column 5 shows identification numbers corresponding to cDNA sequences, coding sequences (exons) predicted from genomic DNA, and/or sequence assemblages comprised of both cDNA and genomic DNA. These sequences were used to assemble the full length polynucleotide sequences of the invention. Columns 6 and 7 of Table 4 show the nucleotide start (5') and stop (3') positions of the cDNA and/or genomic sequences in column 5 relative to their respective full-length sequences.

The identification numbers in Column 5 of Table 4 may refer specifically, for example, to Incyte cDNAs along with their corresponding cDNA libraries. For example, 7246467T8 is the identification number of an Incyte cDNA sequence, and PROSTMY01 is the cDNA library from which it is derived. Incyte cDNAs for which cDNA libraries are not indicated were derived from pooled cDNA libraries (e.g., 71041539V1). Alternatively, the identification numbers in column 5 may refer to GenBank cDNAs or ESTs (e.g., g5745066) which contributed to the assembly of the full length polynucleotide sequences. Alternatively, the identification numbers in column 5 may refer to coding regions predicted by Genscan analysis of genomic DNA. For example, 25 GNN.g7208751\_000002\_002.edit is the identification number of a Genscan-predicted coding sequence, with g7208751 being the GenBank identification number of the sequence to which Genscan was applied. The Genscan-predicted coding sequences may have been edited prior to assembly. (See Example IV.) Alternatively, the identification numbers in column 5 may refer to assemblages of both cDNA and Genscan-predicted exons brought together by an "exon stitching" algorithm. For example, 30 FL1389845\_00001 represents a "stitched" sequence in which 1389845 is the identification number of the cluster of sequences to which the algorithm was applied, and 00001 is the number of the prediction generated by the algorithm. (See Example V.) Alternatively, the identification numbers in column 5 may refer to assemblages of both cDNA and Genscan-predicted exons brought together by an "exon-stretching" algorithm. For example, FL2256251\_g7708357\_000002\_g6103629 is the identification

number of a "stretched" sequence, with 2256251 being the Incyte project identification number, g7708357 being the GenBank identification number of the human genomic sequence to which the "exon-stretching" algorithm was applied, and g6103629 being the GenBank identification number of the nearest GenBank protein homolog. (See Example V.) In some cases, Incyte cDNA coverage 5 redundant with the sequence coverage shown in column 5 was obtained to confirm the final consensus polynucleotide sequence, but the relevant Incyte cDNA identification numbers are not shown.

Table 5 shows the representative cDNA libraries for those full length polynucleotide sequences which were assembled using Incyte cDNA sequences. The representative cDNA library is the Incyte cDNA library which is most frequently represented by the Incyte cDNA sequences which were used to 10 assemble and confirm the above polynucleotide sequences. The tissues and vectors which were used to construct the cDNA libraries shown in Table 5 are described in Table 6.

The invention also encompasses PRTS variants. A preferred PRTS variant is one which has at least about 80%, or alternatively at least about 90%, or even at least about 95% amino acid sequence identity to the PRTS amino acid sequence, and which contains at least one functional or structural 15 characteristic of PRTS.

The invention also encompasses polynucleotides which encode PRTS. In a particular embodiment, the invention encompasses a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:22-42, which encodes PRTS. The polynucleotide sequences of SEQ ID NO:22-42, as presented in the Sequence Listing, embrace the equivalent RNA sequences, 20 wherein occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The invention also encompasses a variant of a polynucleotide sequence encoding PRTS. In particular, such a variant polynucleotide sequence will have at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to the polynucleotide sequence 25 encoding PRTS. A particular aspect of the invention encompasses a variant of a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:22-42 which has at least about 70%, or alternatively at least about 85%, or even at least about 95% polymucleotide sequence identity to a nucleic acid sequence selected from the group consisting of SEQ ID NO:22-42. Any one of the polynucleotide variants described above can encode an amino acid sequence which 30 contains at least one functional or structural characteristic of PRTS.

It will be appreciated by those skilled in the art that as a result of the degeneracy of the genetic code, a multitude of polynucleotide sequences encoding PRTS, some bearing minimal similarity to the polynucleotide sequences of any known and naturally occurring gene, may be produced. Thus, the invention contemplates each and every possible variation of polymucleotide sequence that could be made

by selecting combinations based on possible codon choices. These combinations are made in accordance with the standard triplet genetic code as applied to the polynucleotide sequence of naturally occurring PRTS, and all such variations are to be considered as being specifically disclosed.

Although nucleotide sequences which encode PRTS and its variants are generally capable of hybridizing to the nucleotide sequence of the naturally occurring PRTS under appropriately selected conditions of stringency, it may be advantageous to produce nucleotide sequences encoding PRTS or its derivatives possessing a substantially different codon usage, e.g., inclusion of non-naturally occurring codons. Codons may be selected to increase the rate at which expression of the peptide occurs in a particular prokaryotic or eukaryotic host in accordance with the frequency with which particular codons are utilized by the host. Other reasons for substantially altering the nucleotide sequence encoding PRTS and its derivatives without altering the encoded amino acid sequences include the production of RNA transcripts having more desirable properties, such as a greater half-life, than transcripts produced from the naturally occurring sequence.

The invention also encompasses production of DNA sequences which encode PRTS and PRTS derivatives, or fragments thereof, entirely by synthetic chemistry. After production, the synthetic sequence may be inserted into any of the many available expression vectors and cell systems using reagents well known in the art. Moreover, synthetic chemistry may be used to introduce mutations into a sequence encoding PRTS or any fragment thereof.

Also encompassed by the invention are polynucleotide sequences that are capable of hybridizing to the claimed polynucleotide sequences, and, in particular, to those shown in SEQ ID NO:22-42 and fragments thereof under various conditions of stringency. (See, e.g., Wahl, G.M. and S.L. Berger (1987) Methods Enzymol. 152:399-407; Kimmel, A.R. (1987) Methods Enzymol. 152:507-511.) Hybridization conditions, including annealing and wash conditions, are described in "Definitions."

Methods for DNA sequencing are well known in the art and may be used to practice any of the embodiments of the invention. The methods may employ such enzymes as the Klenow fragment of DNA polymerase I, SEQUENASE (US Biochemical, Cleveland OH), Taq polymerase (Applied Biosystems), thermostable T7 polymerase (Amersham Pharmacia Biotech, Piscataway NJ), or combinations of polymerases and proofreading exonucleases such as those found in the ELONGASE amplification system (Life Technologies, Gaithersburg MD). Preferably, sequence preparation is automated with machines such as the MICROLAB 2200 liquid transfer system (Hamilton, Reno NV), PTC200 thermal cycler (MJ Research, Watertown MA) and ABI CATALYST 800 thermal cycler (Applied Biosystems). Sequencing is then carried out using either the ABI 373 or 377 DNA sequencing system (Applied Biosystems), the MEGABACE 1000 DNA sequencing system (Molecular Dynamics,

Sunnyvale CA), or other systems known in the art. The resulting sequences are analyzed using a variety of algorithms which are well known in the art. (See, e.g., Ausubel, F.M. (1997) Short Protocols in Molecular Biology, John Wiley & Sons, New York NY, unit 7.7; Meyers, R.A. (1995) Molecular Biology and Biotechnology, Wiley VCH, New York NY, pp. 856-853.)

5       The nucleic acid sequences encoding PRTS may be extended utilizing a partial nucleotide sequence and employing various PCR-based methods known in the art to detect upstream sequences, such as promoters and regulatory elements. For example, one method which may be employed, restriction-site PCR, uses universal and nested primers to amplify unknown sequence from genomic DNA within a cloning vector. (See, e.g., Sarkar, G. (1993) *PCR Methods Applic.* 2:318-322.)

10      Another method, inverse PCR, uses primers that extend in divergent directions to amplify unknown sequence from a circularized template. The template is derived from restriction fragments comprising a known genomic locus and surrounding sequences. (See, e.g., Triglia, T. et al. (1988) *Nucleic Acids Res.* 16:8186.) A third method, capture PCR, involves PCR amplification of DNA fragments adjacent to known sequences in human and yeast artificial chromosome DNA. (See, e.g., Lagerstrom, M. et al. 15     (1991) *PCR Methods Applic.* 1:111-119.) In this method, multiple restriction enzyme digestions and ligations may be used to insert an engineered double-stranded sequence into a region of unknown sequence before performing PCR. Other methods which may be used to retrieve unknown sequences are known in the art. (See, e.g., Parker, J.D. et al. (1991) *Nucleic Acids Res.* 19:3055-3060). Additionally, one may use PCR, nested primers, and PROMOTERFINDER libraries (Clontech, Palo 20     Alto CA) to walk genomic DNA. This procedure avoids the need to screen libraries and is useful in finding intron/exon junctions. For all PCR-based methods, primers may be designed using commercially available software, such as OLIGO 4.06 primer analysis software (National Biosciences, Plymouth MN) or another appropriate program, to be about 22 to 30 nucleotides in length; to have a GC content of about 50% or more, and to anneal to the template at temperatures of about 68°C to 25     72°C.

When screening for full length cDNAs, it is preferable to use libraries that have been size-selected to include larger cDNAs. In addition, random-primed libraries, which often include sequences containing the 5' regions of genes, are preferable for situations in which an oligo d(T) library does not yield a full-length cDNA. Genomic libraries may be useful for extension of sequence into 5' 30     non-transcribed regulatory regions.

Capillary electrophoresis systems which are commercially available may be used to analyze the size or confirm the nucleotide sequence of sequencing or PCR products. In particular, capillary sequencing may employ flowable polymers for electrophoretic separation, four different nucleotide-specific, laser-stimulated fluorescent dyes, and a charge coupled device camera for detection of the

emitted wavelengths. Output/light intensity may be converted to electrical signal using appropriate software (e.g., GENOTYPER and SEQUENCE NAVIGATOR, Applied Biosystems), and the entire process from loading of samples to computer analysis and electronic data display may be computer controlled. Capillary electrophoresis is especially preferable for sequencing small DNA fragments  
5 which may be present in limited amounts in a particular sample.

In another embodiment of the invention, polynucleotide sequences or fragments thereof which encode PRTS may be cloned in recombinant DNA molecules that direct expression of PRTS; or fragments or functional equivalents thereof, in appropriate host cells. Due to the inherent degeneracy of the genetic code, other DNA sequences which encode substantially the same or a functionally equivalent  
10 amino acid sequence may be produced and used to express PRTS.

The nucleotide sequences of the present invention can be engineered using methods generally known in the art in order to alter PRTS-encoding sequences for a variety of purposes including, but not limited to, modification of the cloning, processing, and/or expression of the gene product. DNA shuffling by random fragmentation and PCR reassembly of gene fragments and synthetic  
15 oligonucleotides may be used to engineer the nucleotide sequences. For example, oligonucleotide-mediated site-directed mutagenesis may be used to introduce mutations that create new restriction sites, alter glycosylation patterns, change codon preference, produce splice variants, and so forth.

The nucleotides of the present invention may be subjected to DNA shuffling techniques such as MOLECULAR BREEDING (Maxygen Inc., Santa Clara CA; described in U.S. Patent Number  
20 5,837,458; Chang, C.-C. et al. (1999) Nat. Biotechnol. 17:793-797; Christians, F.C. et al. (1999) Nat. Biotechnol. 17:259-264; and Crameri, A. et al. (1996) Nat. Biotechnol. 14:315-319) to alter or improve the biological properties of PRTS, such as its biological or enzymatic activity or its ability to bind to other molecules or compounds. DNA shuffling is a process by which a library of gene variants is produced using PCR-mediated recombination of gene fragments. The library is then  
25 subjected to selection or screening procedures that identify those gene variants with the desired properties. These preferred variants may then be pooled and further subjected to recursive rounds of DNA shuffling and selection/screening. Thus, genetic diversity is created through "artificial" breeding and rapid molecular evolution. For example, fragments of a single gene containing random point mutations may be recombined, screened, and then reshuffled until the desired properties are  
30 optimized. Alternatively, fragments of a given gene may be recombined with fragments of homologous genes in the same gene family, either from the same or different species, thereby maximizing the genetic diversity of multiple naturally occurring genes in a directed and controllable manner.

In another embodiment, sequences encoding PRTS may be synthesized, in whole or in part, using chemical methods well known in the art. (See, e.g., Caruthers, M.H. et al. (1980) Nucleic Acids Symp. Ser. 7:215-223; and Horn, T. et al. (1980) Nucleic Acids Symp. Ser. 7:225-232.) Alternatively, PRTS itself or a fragment thereof may be synthesized using chemical methods. For example, peptide synthesis can be performed using various solution-phase or solid-phase techniques. (See, e.g., Creighton, T. (1984) Proteins, Structures and Molecular Properties, WH Freeman, New York NY, pp. 55-60; and Roberge, J.Y. et al. (1995) Science 269:202-204.) Automated synthesis may be achieved using the ABI 431A peptide synthesizer (Applied Biosystems). Additionally, the amino acid sequence of PRTS, or any part thereof, may be altered during direct synthesis and/or combined with sequences from other proteins, or any part thereof, to produce a variant polypeptide or a polypeptide having a sequence of a naturally occurring polypeptide.

The peptide may be substantially purified by preparative high performance liquid chromatography. (See, e.g., Chiez, R.M. and F.Z. Regnier (1990) Methods Enzymol. 182:392-421.) The composition of the synthetic peptides may be confirmed by amino acid analysis or by sequencing. (See, e.g., Creighton, supra, pp. 28-53.)

In order to express a biologically active PRTS, the nucleotide sequences encoding PRTS or derivatives thereof may be inserted into an appropriate expression vector, i.e., a vector which contains the necessary elements for transcriptional and translational control of the inserted coding sequence in a suitable host. These elements include regulatory sequences, such as enhancers, constitutive and inducible promoters, and 5' and 3' untranslated regions in the vector and in polynucleotide sequences encoding PRTS. Such elements may vary in their strength and specificity. Specific initiation signals may also be used to achieve more efficient translation of sequences encoding PRTS. Such signals include the ATG initiation codon and adjacent sequences, e.g. the Kozak sequence. In cases where sequences encoding PRTS and its initiation codon and upstream regulatory sequences are inserted into the appropriate expression vector, no additional transcriptional or translational control signals may be needed. However, in cases where only coding sequence, or a fragment thereof, is inserted, exogenous translational control signals including an in-frame ATG initiation codon should be provided by the vector. Exogenous translational elements and initiation codons may be of various origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of enhancers appropriate for the particular host cell system used. (See, e.g., Scharf, D. et al. (1994) Results Probl. Cell Differ. 20:125-162.)

Methods which are well known to those skilled in the art may be used to construct expression vectors containing sequences encoding PRTS and appropriate transcriptional and translational control elements. These methods include in vitro recombinant DNA techniques, synthetic techniques, and in

vivo genetic recombination. (See, e.g., Sambrook, J. et al. (1989) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, Plainview NY, ch. 4, 8, and 16-17; Ausubel, F.M. et al. (1995) Current Protocols in Molecular Biology, John Wiley & Sons, New York NY, ch. 9, 13, and 16.)

A variety of expression vector/host systems may be utilized to contain and express sequences encoding PRTS. These include, but are not limited to, microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with viral expression vectors (e.g., baculovirus); plant cell systems transformed with viral expression vectors (e.g., cauliflower mosaic virus, CaMV, or tobacco mosaic virus, TMV) or with bacterial expression vectors (e.g., Ti or pBR322 plasmids); or animal cell systems. (See, e.g., Sambrook, supra; Ausubel, supra; Van Heeke, G. and S.M. Schuster (1989) *J. Biol. Chem.* 264:5503-5509; Engelhard, E.K. et al. (1994) *Proc. Natl. Acad. Sci. USA* 91:3224-3227; Sandig, V. et al. (1996) *Hum. Gene Ther.* 7:1937-1945; Takamatsu, N. (1987) *EMBO J.* 6:307-311; The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196; Logan, J. and T. Shenk (1984) *Proc. Natl. Acad. Sci. USA* 81:3655-3659; and Harrington, J.J. et al. (1997) *Nat. Genet.* 15:345-355.) Expression vectors derived from retroviruses, adenoviruses, or herpes or vaccinia viruses, or from various bacterial plasmids, may be used for delivery of nucleotide sequences to the targeted organ, tissue, or cell population. (See, e.g., Di Nicola, M. et al. (1998) *Cancer Gen. Ther.* 5(6):350-356; Yu, M. et al. (1993) *Proc. Natl. Acad. Sci. USA* 90(13):6340-6344; Buller, R.M. et al. (1985) *Nature* 317(6040):813-815; McGregor, D.P. et al. (1994) *Mol. Immunol.* 31(3):219-226; and Verma, I.M. and N. Somia (1997) *Nature* 389:239-242.) The invention is not limited by the host cell employed.

In bacterial systems, a number of cloning and expression vectors may be selected depending upon the use intended for polynucleotide sequences encoding PRTS. For example, routine cloning, subcloning, and propagation of polynucleotide sequences encoding PRTS can be achieved using a multifunctional *E. coli* vector such as PBLUESCRIPT (Stratagene, La Jolla CA) or PSPORT1 plasmid (Life Technologies). Ligation of sequences encoding PRTS into the vector's multiple cloning site disrupts the *lacZ* gene, allowing a colorimetric screening procedure for identification of transformed bacteria containing recombinant molecules. In addition, these vectors may be useful for in vitro transcription, dideoxy sequencing, single strand rescue with helper phage, and creation of nested deletions in the cloned sequence. (See, e.g., Van Heeke, G. and S.M. Schuster (1989) *J. Biol. Chem.* 264:5503-5509.) When large quantities of PRTS are needed, e.g. for the production of antibodies, vectors which direct high level expression of PRTS may be used. For example, vectors containing the strong, inducible SP6 or T7 bacteriophage promoter may be used.

Yeast expression systems may be used for production of PRTS. A number of vectors containing constitutive or inducible promoters, such as alpha factor, alcohol oxidase, and PGH promoters, may be used in the yeast Saccharomyces cerevisiae or Pichia pastoris. In addition, such vectors direct either the secretion or intracellular retention of expressed proteins and enable integration 5 of foreign sequences into the host genome for stable propagation. (See, e.g., Ausubel, 1995, supra; Bitter, G.A. et al. (1987) Methods Enzymol. 153:516-544; and Scorer, C.A. et al. (1994) Bio/Technology 12:181-184.)

Plant systems may also be used for expression of PRTS. Transcription of sequences encoding PRTS may be driven by viral promoters, e.g., the 35S and 19S promoters of CaMV used alone or in 10 combination with the omega leader sequence from TMV (Takamatsu, N. (1987) EMBO J. 6:307-311). Alternatively, plant promoters such as the small subunit of RUBISCO or heat shock promoters may be used. (See, e.g., Coruzzi, G. et al. (1984) EMBO J. 3:1671-1680; Broglie, R. et al. (1984) Science 224:838-843; and Winter, J. et al. (1991) Results Probl. Cell Differ. 17:85-105.) These constructs can be introduced into plant cells by direct DNA transformation or pathogen-mediated transfection. (See, 15 e.g., The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196.)

In mammalian cells, a number of viral-based expression systems may be utilized. In cases where an adenovirus is used as an expression vector, sequences encoding PRTS may be ligated into an adenovirus transcription/translation complex consisting of the late promoter and tripartite leader 20 sequence. Insertion in a non-essential E1 or E3 region of the viral genome may be used to obtain infective virus which expresses PRTS in host cells. (See, e.g., Logan, J. and T. Shenk (1984) Proc. Natl. Acad. Sci. USA 81:3655-3659.) In addition, transcription enhancers, such as the Rous sarcoma virus (RSV) enhancer, may be used to increase expression in mammalian host cells. SV40 or EBV-based vectors may also be used for high-level protein expression.

25 Human artificial chromosomes (HACs) may also be employed to deliver larger fragments of DNA than can be contained in and expressed from a plasmid. HACs of about 6 kb to 10 Mb are constructed and delivered via conventional delivery methods (liposomes, polycationic amino polymers, or vesicles) for therapeutic purposes. (See, e.g., Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355.)

For long term production of recombinant proteins in mammalian systems, stable expression of 30 PRTS in cell lines is preferred. For example, sequences encoding PRTS can be transformed into cell lines using expression vectors which may contain viral origins of replication and/or endogenous expression elements and a selectable marker gene on the same or on a separate vector. Following the introduction of the vector, cells may be allowed to grow for about 1 to 2 days in enriched media before being switched to selective media. The purpose of the selectable marker is to confer resistance to a

selective agent, and its presence allows growth and recovery of cells which successfully express the introduced sequences. Resistant clones of stably transformed cells may be propagated using tissue culture techniques appropriate to the cell type.

Any number of selection systems may be used to recover transformed cell lines. These include,

5 but are not limited to, the herpes simplex virus thymidine kinase and adenine phosphoribosyltransferase genes, for use in *tk* and *apr* cells, respectively. (See, e.g., Wigler, M. et al. (1977) Cell 11:223-232; Lowy, I. et al. (1980) Cell 22:817-823.) Also, antimetabolite, antibiotic, or herbicide resistance can be used as the basis for selection. For example, *dhfr* confers resistance to methotrexate; *neo* confers resistance to the aminoglycosides neomycin and G-418; and *als* and *pat* confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively. (See, e.g., Wigler, M. et al. (1980) Proc. Natl. Acad. Sci. USA 77:3567-3570; Colbere-Garapin, F. et al. (1981) J. Mol. Biol. 150:1-14.) Additional selectable genes have been described, e.g., *trpB* and *hisD*, which alter cellular requirements for metabolites. (See, e.g., Hartman, S.C. and R.C. Mulligan (1988) Proc. Natl. Acad. Sci. USA 85:8047-8051.) Visible markers, e.g., anthocyanins, green fluorescent proteins (GFP; Clontech),  $\beta$ -glucuronidase and its substrate  $\beta$ -glucuronide, or luciferase and its substrate luciferin may be used.

10 These markers can be used not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system. (See, e.g., Rhodes, C.A. (1995) Methods Mol. Biol. 55:121-131.)

15 Although the presence/absence of marker gene expression suggests that the gene of interest is also present, the presence and expression of the gene may need to be confirmed. For example, if the sequence encoding PRTS is inserted within a marker gene sequence, transformed cells containing sequences encoding PRTS can be identified by the absence of marker gene function. Alternatively, a marker gene can be placed in tandem with a sequence encoding PRTS under the control of a single promoter. Expression of the marker gene in response to induction or selection usually indicates 20 expression of the tandem gene as well.

In general, host cells that contain the nucleic acid sequence encoding PRTS and that express PRTS may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridizations, PCR amplification, and protein bioassay or immunoassay techniques which include membrane, solution, or chip based technologies for the detection and/or quantification of nucleic acid or protein sequences.

30 Immunological methods for detecting and measuring the expression of PRTS using either specific polyclonal or monoclonal antibodies are known in the art. Examples of such techniques include enzyme-linked immunosorbent assays (ELISAs), radioimmunoassays (RIAs), and fluorescence activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing monoclonal

antibodies reactive to two non-interfering epitopes on PRTS is preferred, but a competitive binding assay may be employed. These and other assays are well known in the art. (See, e.g., Hampton, R. et al. (1990) Serological Methods, a Laboratory Manual, APS Press, St. Paul MN, Sect. IV; Coligan, J.E. et al. (1997) Current Protocols in Immunology, Greene Pub. Associates and Wiley-Interscience, New

5 York NY; and Pound, J.D. (1998) Immunochemical Protocols, Humana Press, Totowa NJ.)

A wide variety of labels and conjugation techniques are known by those skilled in the art and may be used in various nucleic acid and amino acid assays. Means for producing labeled hybridization or PCR probes for detecting sequences related to polynucleotides encoding PRTS include oligolabeling, nick translation, end-labeling, or PCR amplification using a labeled nucleotide. Alternatively, the  
10 sequences encoding PRTS, or any fragments thereof, may be cloned into a vector for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by addition of an appropriate RNA polymerase such as T7, T3, or SP6 and labeled nucleotides. These procedures may be conducted using a variety of commercially available kits, such as those provided by Amersham Pharmacia Biotech, Promega (Madison WI), and US  
15 Biochemical. Suitable reporter molecules or labels which may be used for ease of detection include radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents, as well as substrates, cofactors, inhibitors, magnetic particles, and the like.

Host cells transformed with nucleotide sequences encoding PRTS may be cultured under conditions suitable for the expression and recovery of the protein from cell culture. The protein  
20 produced by a transformed cell may be secreted or retained intracellularly depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors containing polynucleotides which encode PRTS may be designed to contain signal sequences which direct secretion of PRTS through a prokaryotic or eukaryotic cell membrane.

In addition, a host cell strain may be chosen for its ability to modulate expression of the  
25 inserted sequences or to process the expressed protein in the desired fashion. Such modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation, and acylation. Post-translational processing which cleaves a "pro" or "pre" form of the protein may also be used to specify protein targeting, folding, and/or activity. Different host cells which have specific cellular machinery and characteristic mechanisms for post-translational activities  
30 (e.g., CHO, HeLa, MDCK, HEK293, and WI38) are available from the American Type Culture Collection (ATCC, Manassas VA) and may be chosen to ensure the correct modification and processing of the foreign protein.

In another embodiment of the invention, natural, modified, or recombinant nucleic acid sequences encoding PRTS may be ligated to a heterologous sequence resulting in translation of a fusion

protein in any of the aforementioned host systems. For example, a chimeric PRTS protein containing a heterologous moiety that can be recognized by a commercially available antibody may facilitate the screening of peptide libraries for inhibitors of PRTS activity. Heterologous protein and peptide moieties may also facilitate purification of fusion proteins using commercially available affinity matrices. Such moieties include, but are not limited to, glutathione S-transferase (GST), maltose binding protein (MBP), thioredoxin (Trx), calmodulin binding peptide (CBP), 6-His, FLAG, *c-myc*, and hemagglutinin (HA). GST, MBP, Trx, CBP, and 6-His enable purification of their cognate fusion proteins on immobilized glutathione, maltose, phenylarsine oxide, calmodulin, and metal-chelate resins, respectively. FLAG, *c-myc*, and hemagglutinin (HA) enable immunoaffinity purification of fusion proteins using commercially available monoclonal and polyclonal antibodies that specifically recognize these epitope tags. A fusion protein may also be engineered to contain a proteolytic cleavage site located between the PRTS encoding sequence and the heterologous protein sequence, so that PRTS may be cleaved away from the heterologous moiety following purification. Methods for fusion protein expression and purification are discussed in Ausubel (1995, *supra*, ch. 10). A variety of commercially available kits may also be used to facilitate expression and purification of fusion proteins.

In a further embodiment of the invention, synthesis of radiolabeled PRTS may be achieved *in vitro* using the TNT rabbit reticulocyte lysate or wheat germ extract system (Promega). These systems couple transcription and translation of protein-coding sequences operably associated with the T7, T3, or SP6 promoters. Translation takes place in the presence of a radiolabeled amino acid precursor, for example,  $^{35}\text{S}$ -methionine.

PRTS of the present invention or fragments thereof may be used to screen for compounds that specifically bind to PRTS. At least one and up to a plurality of test compounds may be screened for specific binding to PRTS. Examples of test compounds include antibodies, oligonucleotides, proteins (e.g., receptors), or small molecules.

In one embodiment, the compound thus identified is closely related to the natural ligand of PRTS, e.g., a ligand or fragment thereof, a natural substrate, a structural or functional mimetic, or a natural binding partner. (See, e.g., Coligan, J.E. et al. (1991) *Current Protocols in Immunology* 1(2): Chapter 5.) Similarly, the compound can be closely related to the natural receptor to which PRTS binds, or to at least a fragment of the receptor, e.g., the ligand binding site. In either case, the compound can be rationally designed using known techniques. In one embodiment, screening for these compounds involves producing appropriate cells which express PRTS, either as a secreted protein or on the cell membrane. Preferred cells include cells from mammals, yeast, *Drosophila*, or *E. coli*. Cells expressing PRTS or cell membrane fractions which contain PRTS are then contacted with

a test compound and binding, stimulation, or inhibition of activity of either PRTS or the compound is analyzed.

An assay may simply test binding of a test compound to the polypeptide, wherein binding is detected by a fluorophore, radioisotope, enzyme conjugate, or other detectable label. For example,

5 the assay may comprise the steps of combining at least one test compound with PRTS, either in solution or affixed to a solid support, and detecting the binding of PRTS to the compound.

Alternatively, the assay may detect or measure binding of a test compound in the presence of a labeled competitor. Additionally, the assay may be carried out using cell-free preparations, chemical libraries, or natural product mixtures, and the test compound(s) may be free in solution or affixed to a

10 solid support.

PRTS of the present invention or fragments thereof may be used to screen for compounds that modulate the activity of PRTS. Such compounds may include agonists, antagonists, or partial or inverse agonists. In one embodiment, an assay is performed under conditions permissive for PRTS activity, wherein PRTS is combined with at least one test compound, and the activity of PRTS in the

15 presence of a test compound is compared with the activity of PRTS in the absence of the test compound. A change in the activity of PRTS in the presence of the test compound is indicative of a compound that modulates the activity of PRTS. Alternatively, a test compound is combined with an in vitro or cell-free system comprising PRTS under conditions suitable for PRTS activity, and the assay is performed. In either of these assays, a test compound which modulates the activity of PRTS may do so

20 indirectly and need not come in direct contact with the test compound. At least one and up to a plurality of test compounds may be screened.

In another embodiment, polynucleotides encoding PRTS or their mammalian homologs may be "knocked out" in an animal model system using homologous recombination in embryonic stem (ES) cells. Such techniques are well known in the art and are useful for the generation of animal models of

25 human disease. (See, e.g., U.S. Patent Number 5,175,383 and U.S. Patent Number 5,767,337.) For example, mouse ES cells, such as the mouse 129/SvJ cell line, are derived from the early mouse embryo and grown in culture. The ES cells are transformed with a vector containing the gene of interest disrupted by a marker gene, e.g., the neomycin phosphotransferase gene (neo; Capecchi, M.R. (1989) Science 244:1288-1292). The vector integrates into the corresponding region of the host genome by

30 homologous recombination. Alternatively, homologous recombination takes place using the Cre-loxP system to knockout a gene of interest in a tissue- or developmental stage-specific manner (Marth, J.D. (1996) Clin. Invest. 97:1999-2002; Wagner, K.U. et al. (1997) Nucleic Acids Res. 25:4323-4330). Transformed ES cells are identified and microinjected into mouse cell blastocysts such as those from the C57BL/6 mouse strain. The blastocysts are surgically transferred to pseudopregnant dams, and the

resulting chimeric progeny are genotyped and bred to produce heterozygous or homozygous strains. Transgenic animals thus generated may be tested with potential therapeutic or toxic agents.

Polynucleotides encoding PRTS may also be manipulated in vitro in ES cells derived from human blastocysts. Human ES cells have the potential to differentiate into at least eight separate cell lineages including endoderm, mesoderm, and ectodermal cell types. These cell lineages differentiate into, for example, neural cells, hematopoietic lineages, and cardiomyocytes (Thomson, J.A. et al. (1998) Science 282:1145-1147).

Polynucleotides encoding PRTS can also be used to create "knockin" humanized animals (pigs) or transgenic animals (mice or rats) to model human disease. With knockin technology, a region of a polynucleotide encoding PRTS is injected into animal ES cells, and the injected sequence integrates into the animal cell genome. Transformed cells are injected into blastulae, and the blastulae are implanted as described above. Transgenic progeny or inbred lines are studied and treated with potential pharmaceutical agents to obtain information on treatment of a human disease. Alternatively, a mammal inbred to overexpress PRTS, e.g., by secreting PRTS in its milk, may also serve as a convenient source of that protein (Janne, J. et al. (1998) Biotechnol. Annu. Rev. 4:55-74).

#### THERAPEUTICS

PRTS are useful for hydrolyzing peptide bonds. Chemical and structural similarity, e.g., in the context of sequences and motifs, exists between regions of PRTS and proteases. In addition, the expression of PRTS is closely associated with hemic, neurological, reproductive, endocrine, urogenital, diseased, teratocarcinoma, and tumorous tissues. Therefore, PRTS appears to play a role in gastrointestinal, cardiovascular, autoimmune/inflammatory, cell proliferative, developmental, epithelial, neurological, and reproductive disorders. In the treatment of disorders associated with increased PRTS expression or activity, it is desirable to decrease the expression or activity of PRTS. In the treatment of disorders associated with decreased PRTS expression or activity, it is desirable to increase the expression or activity of PRTS.

Therefore, in one embodiment, PRTS or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of PRTS. Examples of such disorders include, but are not limited to, a gastrointestinal disorder, such as dysphagia, peptic esophagitis, esophageal spasm, esophageal stricture, esophageal carcinoma, dyspepsia, indigestion, gastritis, gastric carcinoma, anorexia, nausea, emesis, gastroparesis, antral or pyloric edema, abdominal angina, pyrosis, gastroenteritis, intestinal obstruction, infections of the intestinal tract, peptic ulcer, cholelithiasis, cholecystitis, cholestasis, pancreatitis, pancreatic carcinoma, biliary tract disease, hepatitis, hyperbilirubinemia, cirrhosis, passive congestion of the liver, hepatoma, infectious colitis, ulcerative colitis, ulcerative proctitis, Crohn's disease, Whipple's disease, Mallory-

Weiss syndrome, colonic carcinoma, colonic obstruction, irritable bowel syndrome, short bowel syndrome, diarrhea, constipation, gastrointestinal hemorrhage, acquired immunodeficiency syndrome (AIDS) enteropathy, jaundice, hepatic encephalopathy, hepatorenal syndrome, hepatic steatosis, hemochromatosis, Wilson's disease, alpha<sub>1</sub>-antitrypsin deficiency, Reye's syndrome, primary sclerosing

5     cholangitis, liver infarction, portal vein obstruction and thrombosis, centrilobular necrosis, peliosis hepatitis, hepatic vein thrombosis, veno-occlusive disease, preeclampsia, eclampsia, acute fatty liver of pregnancy, intrahepatic cholestasis of pregnancy, and hepatic tumors including nodular hyperplasias, adenomas, and carcinomas; a cardiovascular disorder, such as arteriovenous fistula, atherosclerosis, hypertension, vasculitis, Raynaud's disease, aneurysms, arterial dissections, varicose veins,

10    thrombophlebitis and phlebothrombosis, vascular tumors, and complications of thrombolysis, balloon angioplasty, vascular replacement, and coronary artery bypass graft surgery, congestive heart failure, ischemic heart disease, angina pectoris, myocardial infarction, hypertensive heart disease, degenerative valvular heart disease, calcific aortic valve stenosis, congenitally bicuspid aortic valve, mitral annular calcification, mitral valve prolapse, rheumatic fever and rheumatic heart disease, infective endocarditis,

15    nonbacterial thrombotic endocarditis, endocarditis of systemic lupus erythematosus, carcinoid heart disease, cardiomyopathy, myocarditis, pericarditis, neoplastic heart disease, congenital heart disease, and complications of cardiac transplantation; an autoimmune/inflammatory disorder, such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, atherosclerotic plaque rupture,

20    autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable

25    bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, degradation of articular cartilage, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral,

30    bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; a cell proliferative disorder such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal

gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus; a developmental disorder, such as renal tubular acidosis, anemia, Cushing's syndrome, achondroplastic dwarfism, Duchenne and Becker muscular dystrophy,

5 bone resorption, epilepsy, gonadal dysgenesis, WAGR syndrome (Wilms' tumor, aniridia, genitourinary abnormalities, and mental retardation), Smith-Magenis syndrome, myelodysplastic syndrome, hereditary mucoepithelial dysplasia, hereditary keratodermas, hereditary neuropathies such as Charcot-Marie-Tooth disease and neurofibromatosis, hypothyroidism, hydrocephalus, seizure disorders such as Sydenham's chorea and cerebral palsy, spina bifida, anencephaly, craniorachischisis, congenital

10 glaucoma, cataract, age-related macular degeneration, and sensorineural hearing loss; an epithelial disorder, such as dyshidrotic eczema, allergic contact dermatitis, keratosis pilaris, melasma, vitiligo, actinic keratosis, basal cell carcinoma, squamous cell carcinoma, seborrheic keratosis, folliculitis, herpes simplex, herpes zoster, varicella, candidiasis, dermatophytosis, scabies, insect bites; cherry angioma, keloid, dermatofibroma, acrochordons, urticaria, transient acantholytic dermatosis, xerosis,

15 eczema, atopic dermatitis, contact dermatitis, hand eczema, nummular eczema, lichen simplex chronicus, asteatotic eczema; stasis dermatitis and stasis ulceration, seborrheic dermatitis, psoriasis, lichen planus, pityriasis rosea, impetigo, ecthyma, dermatophytosis, tinea versicolor, warts, acne vulgaris, acne rosacea, pemphigus vulgaris, pemphigus foliaceus, paraneoplastic pemphigus, bullous pemphigoid, herpes gestationis, dermatitis herpetiformis, linear IgA disease, epidermolysis bullosa

20 acquisita, dermatomyositis, lupus erythematosus, scleroderma and morphea, erythroderma, alopecia, figurate skin lesions, telangiectasias, hypopigmentation, hyperpigmentation, vesicles/bullae, exanthems, cutaneous drug reactions, papulonodular skin lesions, chronic non-healing wounds, photosensitivity diseases, epidermolysis bullosa simplex, epidermolytic hyperkeratosis, epidermolytic and nonepidermolytic palmoplantar-keratoderma, ichthyosis bullosa of Siemens, ichthyosis exfoliativa,

25 keratosis palmaris et plantaris, keratosis palmoplantar, palmoplantar keratoderma, keratosis punctata, Meesmann's corneal dystrophy, pachyonychia congenita, white sponge nevus, steatocystoma multiplex, epidermal nevi/epidermolytic hyperkeratosis type, monilethrix, trichothiodystrophy, chronic hepatitis/cryptogenic cirrhosis, and colorectal hyperplasia; a neurological disorder, such as epilepsy, ischemic cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's disease, Pick's disease,

30 Huntington's disease, dementia, Parkinson's disease and other extrapyramidal disorders, amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease, prion diseases including

kuru, Creutzfeldt-Jakob disease, and Gerstmann-Straussler-Scheinker syndrome, fatal familial insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis, encephalotrigeminal syndrome, mental retardation and other developmental disorders of the central nervous system including Down syndrome, cerebral palsy, neuroskeletal disorders, autonomic nervous system disorders, cranial nerve disorders, spinal cord diseases, muscular dystrophy and other neuromuscular disorders, peripheral nervous system disorders, dermatomyositis and polymyositis, inherited, metabolic, endocrine, and toxic myopathies, myasthenia gravis, periodic paralysis, mental disorders including mood, anxiety, and schizophrenic disorders, seasonal affective disorder (SAD), akathesia, amnesia, catatonia, diabetic neuropathy, tardive dyskinesia, dystonias, paranoid psychoses, postherpetic neuralgia, Tourette's disorder, progressive supranuclear palsy, corticobasal degeneration, and familial frontotemporal dementia; and a reproductive disorder, such as infertility, including tubal disease, ovulatory defects, and endometriosis, a disorder of prolactin production, a disruption of the estrous cycle, a disruption of the menstrual cycle, polycystic ovary syndrome, ovarian hyperstimulation syndrome, an endometrial or ovarian tumor, a uterine fibroid, autoimmune disorders, an ectopic pregnancy, and teratogenesis; cancer of the breast, fibrocystic breast disease, and galactorrhea; a disruption of spermatogenesis, abnormal sperm physiology, cancer of the testis, cancer of the prostate, benign prostatic hyperplasia, prostatitis, Peyronie's disease, impotence, carcinoma of the male breast, and gynecomastia.

In another embodiment, a vector capable of expressing PRTS or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of PRTS including, but not limited to, those described above.

In a further embodiment, a composition comprising a substantially purified PRTS in conjunction with a suitable pharmaceutical carrier may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of PRTS including, but not limited to, those provided above.

In still another embodiment, an agonist which modulates the activity of PRTS may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of PRTS including, but not limited to, those listed above.

In a further embodiment, an antagonist of PRTS may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of PRTS. Examples of such disorders include, but are not limited to, those gastrointestinal, cardiovascular, autoimmune/inflammatory, cell proliferative, developmental, epithelial, neurological, and reproductive disorders described above. In one aspect, an antibody which specifically binds PRTS may be used

directly as an antagonist or indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissues which express PRTS.

In an additional embodiment, a vector expressing the complement of the polynucleotide encoding PRTS may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of PRTS including, but not limited to, those described above.

In other embodiments, any of the proteins, antagonists, antibodies, agonists, complementary sequences, or vectors of the invention may be administered in combination with other appropriate therapeutic agents. Selection of the appropriate agents for use in combination therapy may be made by one of ordinary skill in the art, according to conventional pharmaceutical principles. The combination of therapeutic agents may act synergistically to effect the treatment or prevention of the various disorders described above. Using this approach, one may be able to achieve therapeutic efficacy with lower dosages of each agent, thus reducing the potential for adverse side effects.

An antagonist of PRTS may be produced using methods which are generally known in the art. In particular, purified PRTS may be used to produce antibodies or to screen libraries of pharmaceutical agents to identify those which specifically bind PRTS. Antibodies to PRTS may also be generated using methods that are well-known in the art. Such antibodies may include, but are not limited to, polyclonal, monoclonal, chimeric, and single chain antibodies, Fab fragments, and fragments produced by a Fab expression library. Neutralizing antibodies (i.e., those which inhibit dimer formation) are generally preferred for therapeutic use.

For the production of antibodies, various hosts including goats, rabbits, rats, mice, humans, and others may be immunized by injection with PRTS or with any fragment or oligopeptide thereof which has immunogenic properties. Depending on the host species, various adjuvants may be used to increase immunological response. Such adjuvants include, but are not limited to, Freund's, mineral gels such as aluminum hydroxide, and surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, KLH, and dinitrophenol. Among adjuvants used in humans, BCG (bacilli Calmette-Guerin) and Corynebacterium parvum are especially preferable.

It is preferred that the oligopeptides, peptides, or fragments used to induce antibodies to PRTS have an amino acid sequence consisting of at least about 5 amino acids, and generally will consist of at least about 10 amino acids. It is also preferable that these oligopeptides, peptides, or fragments are identical to a portion of the amino acid sequence of the natural protein. Short stretches of PRTS amino acids may be fused with those of another protein, such as KLH, and antibodies to the chimeric molecule may be produced.

Monoclonal antibodies to PRTS may be prepared using any technique which provides for the production of antibody molecules by continuous cell lines in culture. These include, but are not limited

to, the hybridoma technique, the human B-cell hybridoma technique, and the EBV-hybridoma technique. (See, e.g., Kohler, G. et al. (1975) *Nature* 256:495-497; Kozbor, D. et al. (1985) *J. Immunol. Methods* 81:31-42; Cote, R.J. et al. (1983) *Proc. Natl. Acad. Sci. USA* 80:2026-2030; and Cole, S.P. et al. (1984) *Mol. Cell Biol.* 62:109-120.)

5 In addition, techniques developed for the production of "chimeric antibodies," such as the splicing of mouse antibody genes to human antibody genes to obtain a molecule with appropriate antigen specificity and biological activity, can be used. (See, e.g., Morrison, S.L. et al. (1984) *Proc. Natl. Acad. Sci. USA* 81:6851-6855; Neuberger, M.S. et al. (1984) *Nature* 312:604-608; and Takeda, S. et al. (1985) *Nature* 314:452-454.) Alternatively, techniques described for the production of single 10 chain antibodies may be adapted, using methods known in the art, to produce PRTS-specific single chain antibodies. Antibodies with related specificity, but of distinct idiotypic composition, may be generated by chain shuffling from random combinatorial immunoglobulin libraries. (See, e.g., Burton, D.R. (1991) *Proc. Natl. Acad. Sci. USA* 88:10134-10137.)

15 Antibodies may also be produced by inducing *in vivo* production in the lymphocyte population or by screening immunoglobulin libraries or panels of highly specific binding reagents as disclosed in the literature. (See, e.g.; Orlandi, R. et al. (1989) *Proc. Natl. Acad. Sci. USA* 86:3833-3837; Winter, G. et al. (1991) *Nature* 349:293-299.)

20 Antibody fragments which contain specific binding sites for PRTS may also be generated. For example, such fragments include, but are not limited to, F(ab')<sub>2</sub> fragments produced by pepsin digestion 20 of the antibody molecule and Fab fragments generated by reducing the disulfide bridges of the F(ab')<sub>2</sub> fragments. Alternatively, Fab expression libraries may be constructed to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity. (See, e.g., Huse, W.D. et al. (1989) *Science* 246:1275-1281.)

25 Various immunoassays may be used for screening to identify antibodies having the desired specificity. Numerous protocols for competitive binding or immunoradiometric assays using either polyclonal or monoclonal antibodies with established specificities are well known in the art. Such immunoassays typically involve the measurement of complex formation between PRTS and its specific antibody. A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering PRTS epitopes is generally used, but a competitive binding assay may also be employed 30 (Pound, supra).

Various methods such as Scatchard analysis in conjunction with radioimmunoassay techniques may be used to assess the affinity of antibodies for PRTS. Affinity is expressed as an association constant, K<sub>a</sub>, which is defined as the molar concentration of PRTS-antibody complex divided by the molar concentrations of free antigen and free antibody under equilibrium conditions. The K<sub>a</sub> determined

for a preparation of polyclonal antibodies, which are heterogeneous in their affinities for multiple PRTS epitopes, represents the average affinity, or avidity, of the antibodies for PRTS. The  $K_a$  determined for a preparation of monoclonal antibodies, which are monospecific for a particular PRTS epitope, represents a true measure of affinity. High-affinity antibody preparations with  $K_a$  ranging from about 5  $10^9$  to  $10^{12}$  L/mole are preferred for use in immunoassays in which the PRTS-antibody complex must withstand rigorous manipulations. Low-affinity antibody preparations with  $K_a$  ranging from about  $10^6$  to  $10^7$  L/mole are preferred for use in immunopurification and similar procedures which ultimately require dissociation of PRTS, preferably in active form, from the antibody (Catty, D. (1988) Antibodies, Volume I: A Practical Approach, IRL Press, Washington DC; Liddell, J.E. and A. Cryer (1991) A Practical Guide to Monoclonal Antibodies, John Wiley & Sons, New York NY).

The titer and avidity of polyclonal antibody preparations may be further evaluated to determine the quality and suitability of such preparations for certain downstream applications. For example, a polyclonal antibody preparation containing at least 1-2 mg specific antibody/ml, preferably 5-10 mg specific antibody/ml, is generally employed in procedures requiring precipitation of PRTS-antibody complexes. Procedures for evaluating antibody specificity, titer, and avidity, and guidelines for antibody quality and usage in various applications, are generally available. (See, e.g., Catty, supra, and Coligan et al., supra.)

In another embodiment of the invention, the polynucleotides encoding PRTS, or any fragment or complement thereof, may be used for therapeutic purposes. In one aspect, modifications of gene expression can be achieved by designing complementary sequences or antisense molecules (DNA, RNA, PNA, or modified oligonucleotides) to the coding or regulatory regions of the gene encoding PRTS. Such technology is well known in the art, and antisense oligonucleotides or larger fragments can be designed from various locations along the coding or control regions of sequences encoding PRTS. (See, e.g., Agrawal, S., ed. (1996) Antisense Therapeutics, Humana Press Inc., Totowa NJ.)

25 In therapeutic use, any gene delivery system suitable for introduction of the antisense sequences into appropriate target cells can be used. Antisense sequences can be delivered intracellularly in the form of an expression plasmid which, upon transcription, produces a sequence complementary to at least a portion of the cellular sequence encoding the target protein. (See, e.g., Slater, J.E. et al. (1998) *J. Allergy Cli. Immunol.* 102(3):469-475; and Scanlon, K.J. et al. (1995) 30 9(13):1288-1296.) Antisense sequences can also be introduced intracellularly through the use of viral vectors, such as retrovirus and adeno-associated virus vectors. (See, e.g., Miller, A.D. (1990) *Blood* 76:271; Ausubel, supra; Uckert, W. and W. Walther (1994) *Pharmacol. Ther.* 63(3):323-347.) Other gene delivery mechanisms include liposome-derived systems, artificial viral envelopes, and other systems known in the art. (See, e.g., Rossi, J.J. (1995) *Br. Med. Bull.* 51(1):217-225; Boado, R.J. et

al. (1998) J. Pharm. Sci. 87(11):1308-1315; and Morris, M.C. et al. (1997) Nucleic Acids Res. 25(14):2730-2736.)

In another embodiment of the invention, polynucleotides encoding PRTS may be used for somatic or germline gene therapy. Gene therapy may be performed to (i) correct a genetic deficiency (e.g., in the cases of severe combined immunodeficiency (SCID)-X1 disease characterized by X-linked inheritance (Cavazzana-Calvo, M. et al. (2000) Science 288:669-672), severe combined immunodeficiency syndrome associated with an inherited adenosine deaminase (ADA) deficiency (Blaese, R.M. et al. (1995) Science 270:475-480; Bordignon, C. et al. (1995) Science 270:470-475), cystic fibrosis (Zabner, J. et al. (1993) Cell 75:207-216; Crystal, R.G. et al. (1995) Hum. Gene Therapy 6:643-666; Crystal, R.G. et al. (1995) Hum. Gene Therapy 6:667-703), thalassamias, familial hypercholesterolemia, and hemophilia resulting from Factor VIII or Factor IX deficiencies (Crystal, R.G. (1995) Science 270:404-410; Verma, I.M. and N. Somia (1997) Nature 389:239-242)), (ii) express a conditionally lethal gene product (e.g., in the case of cancers which result from unregulated cell proliferation), or (iii) express a protein which affords protection against intracellular parasites (e.g., 5 against human retroviruses, such as human immunodeficiency virus (HIV) (Baltimore, D. (1988) Nature 335:395-396; Poeschla, E. et al. (1996) Proc. Natl. Acad. Sci. USA. 93:11395-11399), hepatitis B or C virus (HBV, HCV); fungal parasites, such as Candida albicans and Paracoccidioides brasiliensis; and protozoan parasites such as Plasmodium falciparum and Trypanosoma cruzi). In the case where a genetic deficiency in PRTS expression or regulation causes disease, the expression of 10 PRTS from an appropriate population of transduced cells may alleviate the clinical manifestations caused by the genetic deficiency.

In a further embodiment of the invention, diseases or disorders caused by deficiencies in PRTS are treated by constructing mammalian expression vectors encoding PRTS and introducing these vectors by mechanical means into PRTS-deficient cells. Mechanical transfer technologies for use with 15 cells in vivo or ex vitro include (i) direct DNA microinjection into individual cells, (ii) ballistic gold particle delivery, (iii) liposome-mediated transfection, (iv) receptor-mediated gene transfer, and (v) the use of DNA transposons (Morgan, R.A. and W.F. Anderson (1993) Annu. Rev. Biochem. 62:191-217; Ivics, Z. (1997) Cell 91:501-510; Boulay, J-L. and H. Récipon (1998) Curr. Opin. Biotechnol. 9:445-450).

20 Expression vectors that may be effective for the expression of PRTS include, but are not limited to, the PCDNA 3.1, EPITAG, PRCCMV2, PREP, PVAX vectors (Invitrogen, Carlsbad CA), PCMV-SCRIPT, PCMV-TAG, PEGSH/PERV (Stratagene, La Jolla CA), and PTET-OFF, PTET-ON, PTRE2, PTRE2-LUC, PTK-HYG (Clontech, Palo Alto CA). PRTS may be expressed using (i) a constitutively active promoter, (e.g., from cytomegalovirus (CMV), Rous sarcoma virus

(RSV), SV40 virus, thymidine kinase (TK), or  $\beta$ -actin genes), (ii) an inducible promoter (e.g., the tetracycline-regulated promoter (Gossen, M. and H. Bujard (1992) Proc. Natl. Acad. Sci. USA 89:5547-5551; Gossen, M. et al. (1995) Science 268:1766-1769; Rossi, F.M.V. and H.M. Blau (1998) Curr. Opin. Biotechnol. 9:451-456), commercially available in the T-REX plasmid (Invitrogen)); the 5 ecdisone-inducible promoter (available in the plasmids PVGRXR and PIND; Invitrogen); the FK506/rapamycin inducible promoter; or the RU486/mifepristone inducible promoter (Rossi, F.M.V. and Blau, H.M. *supra*), or (iii) a tissue-specific promoter or the native promoter of the endogenous gene encoding PRTS from a normal individual.

Commercially available liposome transformation kits (e.g., the PERFECT LIPID 10 TRANSFECTION KIT, available from Invitrogen) allow one with ordinary skill in the art to deliver polynucleotides to target cells in culture and require minimal effort to optimize experimental parameters. In the alternative, transformation is performed using the calcium phosphate method (Graham, F.L. and A.J. Eb (1973) Virology 52:456-467), or by electroporation (Neumann, E. et al. (1982) EMBO J. 1:841-845). The introduction of DNA to primary cells requires modification of these 15 standardized mammalian transfection protocols.

In another embodiment of the invention, diseases or disorders caused by genetic defects with respect to PRTS expression are treated by constructing a retrovirus vector consisting of (i) the polynucleotide encoding PRTS under the control of an independent promoter or the retrovirus long terminal repeat (LTR)-promoter; (ii) appropriate RNA packaging signals, and (iii) a Rev-responsive element (RRE) along with additional retrovirus *cis*-acting RNA sequences and coding sequences required for efficient vector propagation. Retrovirus vectors (e.g., PFB and PFBNEO) are commercially available (Stratagene) and are based on published data (Riviere, I. et al. (1995) Proc. Natl. Acad. Sci. USA 92:6733-6737), incorporated by reference herein. The vector is propagated in an appropriate vector producing cell line (VPCL) that expresses an envelope gene with a tropism for 20 receptors on the target cells or a promiscuous envelope protein such as VSVg (Armentano, D. et al. (1987) J. Virol. 61:1647-1650; Bender, M.A. et al. (1987) J. Virol. 61:1639-1646; Adam, M.A. and A.D. Miller (1988) J. Virol. 62:3802-3806; Dull, T. et al. (1998) J. Virol. 72:8463-8471; Zufferey, R. et al. (1998) J. Virol. 72:9873-9880). U.S. Patent Number 5,910,434 to Rigg ("Method for obtaining retrovirus packaging cell lines producing high transducing efficiency retroviral supernatant") discloses a 25 method for obtaining retrovirus packaging cell lines and is hereby incorporated by reference. Propagation of retrovirus vectors, transduction of a population of cells (e.g., CD4 $^{+}$  T-cells), and the return of transduced cells to a patient are procedures well known to persons skilled in the art of gene therapy and have been well documented (Ranga, U. et al. (1997) J. Virol. 71:7020-7029; Bauer, G. et 30

al. (1997) Blood 89:2259-2267; Bonyhadi, M.L. (1997) J. Virol. 71:4707-4716; Ranga, U. et al. (1998) Proc. Natl. Acad. Sci. USA 95:1201-1206; Su, L. (1997) Blood 89:2283-2290).

In the alternative, an adenovirus-based gene therapy delivery system is used to deliver polynucleotides encoding PRTS to cells which have one or more genetic abnormalities with respect to the expression of PRTS. The construction and packaging of adenovirus-based vectors are well known to those with ordinary skill in the art. Replication defective adenovirus vectors have proven to be versatile for importing genes encoding immunoregulatory proteins into intact islets in the pancreas (Csete, M.E. et al. (1995) Transplantation 27:263-268). Potentially useful adenoviral vectors are described in U.S. Patent Number 5,707,618 to Armentano ("Adenovirus vectors for gene therapy"), hereby incorporated by reference. For adenoviral vectors, see also Antinozzi, P.A. et al. (1999) Annu. Rev. Nutr. 19:511-544 and Verma, I.M. and N. Somia (1997) Nature 389:239-242, both incorporated by reference herein.

In another alternative, a herpes-based, gene therapy delivery system is used to deliver polynucleotides encoding PRTS to target cells which have one or more genetic abnormalities with respect to the expression of PRTS. The use of herpes simplex virus (HSV)-based vectors may be especially valuable for introducing PRTS to cells of the central nervous system, for which HSV has a tropism. The construction and packaging of herpes-based vectors are well known to those with ordinary skill in the art. A replication-competent herpes simplex virus (HSV) type 1-based vector has been used to deliver a reporter gene to the eyes of primates (Liu, X. et al. (1999) Exp. Eye Res. 69:385-395). The construction of a HSV-1 virus vector has also been disclosed in detail in U.S. Patent Number 5,804,413 to DeLuca ("Herpes simplex virus strains for gene transfer"), which is hereby incorporated by reference. U.S. Patent Number 5,804,413 teaches the use of recombinant HSV d92 which consists of a genome containing at least one exogenous gene to be transferred to a cell under the control of the appropriate promoter for purposes including human gene therapy. Also taught by this patent are the construction and use of recombinant HSV strains deleted for ICP4, ICP27 and ICP22. For HSV vectors, see also Goins, W.F. et al. (1999) J. Virol. 73:519-532 and Xu, H. et al. (1994) Dev. Biol. 163:152-161, hereby incorporated by reference. The manipulation of cloned herpesvirus sequences, the generation of recombinant virus following the transfection of multiple plasmids containing different segments of the large herpesvirus genomes, the growth and propagation of herpesvirus, and the infection of cells with herpesvirus are techniques well known to those of ordinary skill in the art.

In another alternative, an alphavirus (positive, single-stranded RNA virus) vector is used to deliver polynucleotides encoding PRTS to target cells. The biology of the prototypic alphavirus, Semliki Forest Virus (SFV), has been studied extensively and gene transfer vectors have been based on

the SFV genome (Garoff, H. and K.-J. Li (1998) Curr. Opin. Biotechnol. 9:464-469). During alphavirus RNA replication, a subgenomic RNA is generated that normally encodes the viral capsid proteins. This subgenomic RNA replicates to higher levels than the full length genomic RNA, resulting in the overproduction of capsid proteins relative to the viral proteins with enzymatic activity (e.g., 5 protease and polymerase). Similarly, inserting the coding sequence for PRTS into the alphavirus genome in place of the capsid-coding region results in the production of a large number of PRTS-coding RNAs and the synthesis of high levels of PRTS in vector transduced cells. While alphavirus infection is typically associated with cell lysis within a few days, the ability to establish a persistent infection in hamster normal kidney cells (BHK-21) with a variant of Sindbis virus (SIN) indicates that the lytic 10 replication of alphaviruses can be altered to suit the needs of the gene therapy application (Dryga, S.A. et al. (1997) Virology 228:74-83). The wide host range of alphaviruses will allow the introduction of PRTS into a variety of cell types. The specific transduction of a subset of cells in a population may require the sorting of cells prior to transduction. The methods of manipulating infectious cDNA clones 15 of alphaviruses, performing alphavirus cDNA and RNA transfections, and performing alphavirus infections, are well known to those with ordinary skill in the art.

Oligonucleotides derived from the transcription initiation site, e.g., between about positions -10 and +10 from the start site, may also be employed to inhibit gene expression. Similarly, inhibition can be achieved using triple helix base-pairing methodology. Triple helix pairing is useful because it causes inhibition of the ability of the double helix to open sufficiently for the binding of polymerases, 20 transcription factors, or regulatory molecules. Recent therapeutic advances using triplex DNA have been described in the literature. (See, e.g., Gee, J.E. et al. (1994) in Huber, B.E. and B.I. Carr, Molecular and Immunologic Approaches, Futura Publishing, Mt. Kisco NY, pp. 163-177.) A complementary sequence or antisense molecule may also be designed to block translation of mRNA by preventing the transcript from binding to ribosomes.

25 Ribozymes, enzymatic RNA molecules, may also be used to catalyze the specific cleavage of RNA. The mechanism of ribozyme action involves sequence-specific hybridization of the ribozyme molecule to complementary target RNA, followed by endonucleolytic cleavage. For example, engineered hammerhead motif ribozyme molecules may specifically and efficiently catalyze endonucleolytic cleavage of sequences encoding PRTS.

30 Specific ribozyme cleavage sites within any potential RNA target are initially identified by scanning the target molecule for ribozyme cleavage sites, including the following sequences: GUA, GUU, and GUC. Once identified, short RNA sequences of between 15 and 20 ribonucleotides, corresponding to the region of the target gene containing the cleavage site, may be evaluated for secondary structural features which may render the oligonucleotide inoperable. The suitability of

candidate targets may also be evaluated by testing accessibility to hybridization with complementary oligonucleotides using ribonuclease protection assays.

Complementary ribonucleic acid molecules and ribozymes of the invention may be prepared by any method known in the art for the synthesis of nucleic acid molecules. These include techniques for

5 chemically synthesizing oligonucleotides such as solid phase phosphoramidite chemical synthesis.

Alternatively, RNA molecules may be generated by *in vitro* and *in vivo* transcription of DNA sequences encoding PRTS. Such DNA sequences may be incorporated into a wide variety of vectors with suitable RNA polymerase promoters such as T7 or SP6. Alternatively, these cDNA constructs that synthesize complementary RNA, constitutively or inducibly, can be introduced into cell lines, cells, or tissues.

10 RNA molecules may be modified to increase intracellular stability and half-life. Possible modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3' ends of the molecule, or the use of phosphorothioate or 2' O-methyl rather than phosphodiesterate linkages within the backbone of the molecule. This concept is inherent in the production of PNAs and can be extended in all of these molecules by the inclusion of nontraditional bases such as inosine, queosine, and

15 wybutosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytidine, guanine, thymine, and uridine which are not as easily recognized by endogenous endonucleases.

An additional embodiment of the invention encompasses a method for screening for a compound which is effective in altering expression of a polynucleotide encoding PRTS. Compounds which may be effective in altering expression of a specific polynucleotide may include, but are not limited to, oligonucleotides, antisense oligonucleotides, triple helix-forming oligonucleotides, transcription factors and other polypeptide transcriptional regulators, and non-macromolecular chemical entities which are capable of interacting with specific polynucleotide sequences. Effective compounds may alter polynucleotide expression by acting as either inhibitors or promoters of polynucleotide expression. Thus, in the treatment of disorders associated with increased PRTS expression or activity, a compound which specifically inhibits expression of the polynucleotide encoding PRTS may be therapeutically useful, and in the treatment of disorders associated with decreased PRTS expression or activity, a compound which specifically promotes expression of the polynucleotide encoding PRTS may be therapeutically useful.

At least one, and up to a plurality, of test compounds may be screened for effectiveness in

30 altering expression of a specific polynucleotide. A test compound may be obtained by any method commonly known in the art, including chemical modification of a compound known to be effective in altering polynucleotide expression; selection from an existing, commercially-available or proprietary library of naturally-occurring or non-natural chemical compounds; rational design of a compound based on chemical and/or structural properties of the target polynucleotide; and selection from a

library of chemical compounds created combinatorially or randomly. A sample comprising a polynucleotide encoding PRTS is exposed to at least one test compound thus obtained. The sample may comprise, for example, an intact or permeabilized cell, or an in vitro cell-free or reconstituted biochemical system. Alterations in the expression of a polynucleotide encoding PRTS are assayed by

5 any method commonly known in the art. Typically, the expression of a specific nucleotide is detected by hybridization with a probe having a nucleotide sequence complementary to the sequence of the polynucleotide encoding PRTS. The amount of hybridization may be quantified, thus forming the basis for a comparison of the expression of the polynucleotide both with and without exposure to one or more test compounds. Detection of a change in the expression of a polynucleotide exposed to a

10 test compound indicates that the test compound is effective in altering the expression of the polynucleotide. A screen for a compound effective in altering expression of a specific polynucleotide can be carried out, for example, using a Schizosaccharomyces pombe gene expression system (Atkins, D. et al. (1999) U.S. Patent No. 5,932,435; Arndt, G.M. et al. (2000) Nucleic Acids Res. 28:E15) or a human cell line such as HeLa cell (Clarke, M.L. et al. (2000) Biochem. Biophys. Res. Commun. 268:8-13). A particular embodiment of the present invention involves screening a combinatorial library of oligonucleotides (such as deoxyribonucleotides, ribonucleotides, peptide nucleic acids, and modified oligonucleotides) for antisense activity against a specific polynucleotide sequence (Bruice, T.W. et al. (1997) U.S. Patent No. 5,686,242; Bruice, T.W. et al. (2000) U.S. Patent No. 6,022,691).

15 Many methods for introducing vectors into cells or tissues are available and equally suitable for use in vivo, in vitro, and ex vivo. For ex vivo therapy, vectors may be introduced into stem cells taken from the patient and clonally propagated for autologous transplant back into that same patient.

20 Delivery by transfection, by liposome injections, or by polycationic amino polymers may be achieved using methods which are well known in the art. (See, e.g., Goldman, C.K. et al. (1997) Nat. Biotechnol. 15:462-466.)

25 Any of the therapeutic methods described above may be applied to any subject in need of such therapy, including, for example, mammals such as humans, dogs, cats, cows, horses, rabbits, and monkeys.

An additional embodiment of the invention relates to the administration of a composition which generally comprises an active ingredient formulated with a pharmaceutically acceptable excipient.

30 Excipients may include, for example, sugars, starches, celluloses, gums, and proteins. Various formulations are commonly known and are thoroughly discussed in the latest edition of Remington's Pharmaceutical Sciences (Maack Publishing, Easton PA). Such compositions may consist of PRTS, antibodies to PRTS, and mimetics, agonists, antagonists, or inhibitors of PRTS.

The compositions utilized in this invention may be administered by any number of routes including, but not limited to, oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, pulmonary, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, or rectal means.

5 Compositions for pulmonary administration may be prepared in liquid or dry powder form. These compositions are generally aerosolized immediately prior to inhalation by the patient. In the case of small molecules (e.g. traditional low molecular weight organic drugs), aerosol delivery of fast-acting formulations is well-known in the art. In the case of macromolecules (e.g. larger peptides and proteins), recent developments in the field of pulmonary delivery via the alveolar region of the lung have enabled  
10 the practical delivery of drugs such as insulin to blood circulation (see, e.g., Patton, J.S. et al., U.S. Patent No. 5,997,848). Pulmonary delivery has the advantage of administration without needle injection, and obviates the need for potentially toxic penetration enhancers.

Compositions suitable for use in the invention include compositions wherein the active ingredients are contained in an effective amount to achieve the intended purpose. The determination of  
15 an effective dose is well within the capability of those skilled in the art.

Specialized forms of compositions may be prepared for direct intracellular delivery of macromolecules comprising PRTS or fragments thereof. For example, liposome preparations containing a cell-impermeable macromolecule may promote cell fusion and intracellular delivery of the macromolecule. Alternatively, PRTS or a fragment thereof may be joined to a short cationic N-  
20 terminal portion from the HIV Tat-1 protein. Fusion proteins thus generated have been found to transduce into the cells of all tissues, including the brain, in a mouse model system (Schwarze, S.R. et al. (1999) Science 285:1569-1572).

For any compound, the therapeutically effective dose can be estimated initially either in cell culture assays, e.g., of neoplastic cells, or in animal models such as mice, rats, rabbits, dogs, monkeys,  
25 or pigs. An animal model may also be used to determine the appropriate concentration range and route of administration. Such information can then be used to determine useful doses and routes for administration in humans.

A therapeutically effective dose refers to that amount of active ingredient, for example PRTS or fragments thereof, antibodies of PRTS, and agonists, antagonists or inhibitors of PRTS, which  
30 ameliorates the symptoms or condition. Therapeutic efficacy and toxicity may be determined by standard pharmaceutical procedures in cell cultures or with experimental animals, such as by calculating the ED<sub>50</sub> (the dose therapeutically effective in 50% of the population) or LD<sub>50</sub> (the dose lethal to 50% of the population) statistics. The dose ratio of toxic to therapeutic effects is the therapeutic index, which can be expressed as the LD<sub>50</sub>/ED<sub>50</sub> ratio. Compositions which exhibit large

therapeutic indices are preferred. The data obtained from cell culture assays and animal studies are used to formulate a range of dosage for human use. The dosage contained in such compositions is preferably within a range of circulating concentrations that includes the ED<sub>50</sub> with little or no toxicity. The dosage varies within this range depending upon the dosage form employed, the sensitivity of the patient, and the route of administration.

The exact dosage will be determined by the practitioner, in light of factors related to the subject requiring treatment. Dosage and administration are adjusted to provide sufficient levels of the active moiety or to maintain the desired effect. Factors which may be taken into account include the severity of the disease state, the general health of the subject, the age, weight, and gender of the subject, time and frequency of administration, drug combination(s), reaction sensitivities, and response to therapy. Long-acting compositions may be administered every 3 to 4 days, every week, or biweekly depending on the half-life and clearance rate of the particular formulation.

Normal dosage amounts may vary from about 0.1 µg to 100,000 µg, up to a total dose of about 1 gram, depending upon the route of administration. Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art. Those skilled in the art will employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

#### DIAGNOSTICS

In another embodiment, antibodies which specifically bind PRTS may be used for the diagnosis of disorders characterized by expression of PRTS, or in assays to monitor patients being treated with PRTS or agonists, antagonists, or inhibitors of PRTS. Antibodies useful for diagnostic purposes may be prepared in the same manner as described above for therapeutics. Diagnostic assays for PRTS include methods which utilize the antibody and a label to detect PRTS in human body fluids or in extracts of cells or tissues. The antibodies may be used with or without modification, and may be labeled by covalent or non-covalent attachment of a reporter molecule. A wide variety of reporter molecules, several of which are described above, are known in the art and may be used.

A variety of protocols for measuring PRTS, including ELISAs, RIAs, and FACS, are known in the art and provide a basis for diagnosing altered or abnormal levels of PRTS expression. Normal or standard values for PRTS expression are established by combining body fluids or cell extracts taken from normal mammalian subjects, for example, human subjects, with antibodies to PRTS under conditions suitable for complex formation. The amount of standard complex formation may be quantitated by various methods, such as photometric means. Quantities of PRTS expressed in subject,

control, and disease samples from biopsied tissues are compared with the standard values. Deviation between standard and subject values establishes the parameters for diagnosing disease.

In another embodiment of the invention, the polynucleotides encoding PRTS may be used for diagnostic purposes. The polynucleotides which may be used include oligonucleotide sequences, 5 complementary RNA and DNA molecules, and PNAs. The polynucleotides may be used to detect and quantify gene expression in biopsied tissues in which expression of PRTS may be correlated with disease. The diagnostic assay may be used to determine absence, presence, and excess expression of PRTS, and to monitor regulation of PRTS levels during therapeutic intervention.

In one aspect, hybridization with PCR probes which are capable of detecting polynucleotide 10 sequences, including genomic sequences, encoding PRTS or closely related molecules may be used to identify nucleic acid sequences which encode PRTS. The specificity of the probe, whether it is made from a highly specific region, e.g., the 5' regulatory region, or from a less specific region, e.g., a conserved motif, and the stringency of the hybridization or amplification will determine whether the probe identifies only naturally occurring sequences encoding PRTS, allelic variants, or related 15 sequences.

Probes may also be used for the detection of related sequences, and may have at least 50% sequence identity to any of the PRTS encoding sequences. The hybridization probes of the subject invention may be DNA or RNA and may be derived from the sequence of SEQ ID NO:22-42 or from genomic sequences including promoters, enhancers, and introns of the PRTS gene.

20 Means for producing specific hybridization probes for DNAs encoding PRTS include the cloning of polynucleotide sequences encoding PRTS or PRTS derivatives into vectors for the production of mRNA probes. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes *in vitro* by means of the addition of the appropriate RNA polymerases and the appropriate labeled nucleotides. Hybridization probes may be labeled by a variety 25 of reporter groups, for example, by radionuclides such as  $^{32}\text{P}$  or  $^{35}\text{S}$ , or by enzymatic labels, such as alkaline phosphatase coupled to the probe via avidin/biotin coupling systems, and the like.

Polynucleotide sequences encoding PRTS may be used for the diagnosis of disorders associated 30 with expression of PRTS. Examples of such disorders include, but are not limited to, a gastrointestinal disorder, such as dysphagia, peptic esophagitis, esophageal spasm, esophageal stricture, esophageal carcinoma, dyspepsia, indigestion, gastritis, gastric carcinoma, anorexia, nausea, emesis, gastroparesis, antral or pyloric edema, abdominal angina, pyrosis, gastroenteritis, intestinal obstruction, infections of the intestinal tract, peptic ulcer, cholelithiasis, cholecystitis, cholestasis, pancreatitis, pancreatic carcinoma, biliary tract disease, hepatitis, hyperbilirubinemia, cirrhosis, passive congestion of the liver, hepatoma, infectious colitis, ulcerative colitis, ulcerative proctitis, Crohn's disease, Whipple's disease,

Mallory-Weiss syndrome, colonic carcinoma, colonic obstruction, irritable bowel syndrome, short bowel syndrome, diarrhea, constipation, gastrointestinal hemorrhage, acquired immunodeficiency syndrome (AIDS) enteropathy, jaundice, hepatic encephalopathy, hepatorenal syndrome, hepatic steatosis, hemochromatosis, Wilson's disease, alpha<sub>1</sub>-antitrypsin deficiency, Reye's syndrome, primary

5 sclerosing cholangitis, liver infarction, portal vein obstruction and thrombosis, centrilobular necrosis, peliosis hepatitis, hepatic vein thrombosis, veno-occlusive disease, preeclampsia, eclampsia, acute fatty liver of pregnancy, intrahepatic cholestasis of pregnancy, and hepatic tumors including nodular hyperplasias, adenomas, and carcinomas; a cardiovascular disorder, such as arteriovenous fistula, atherosclerosis, hypertension, vasculitis, Raynaud's disease, aneurysms, arterial dissections, varicose

10 veins, thrombophlebitis and phlebothrombosis, vascular tumors, and complications of thrombolysis, balloon angioplasty, vascular replacement, and coronary artery bypass graft surgery, congestive heart failure, ischemic heart disease, angina pectoris, myocardial infarction, hypertensive heart disease, degenerative valvular heart disease, calcific aortic valve stenosis, congenitally bicuspid aortic valve, mitral annular calcification, mitral valve prolapse, rheumatic fever and rheumatic heart disease,

15 infective endocarditis, nonbacterial thrombotic endocarditis, endocarditis of systemic lupus erythematosus, carcinoid heart disease, cardiomyopathy, myocarditis, pericarditis, neoplastic heart disease, congenital heart disease, and complications of cardiac transplantation; an autoimmune/inflammatory disorder, such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia,

20 asthma, atherosclerosis, atherosclerotic plaque rupture, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's

25 thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, degradation of articular cartilage, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and

30 extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; a cell proliferative disorder such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in

particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus; a developmental disorder, such as renal tubular acidosis, anemia, Cushing's syndrome, achondroplastic dwarfism, Duchenne and Becker muscular dystrophy, bone resorption, epilepsy, gonadal dysgenesis, WAGR syndrome (Wilms' tumor, aniridia, genitourinary abnormalities, and mental retardation), Smith-Magenis syndrome, myelodysplastic syndrome, hereditary mucoepithelial dysplasia, hereditary keratodermas, hereditary neuropathies such as Charcot-Marie-Tooth disease and neurofibromatosis, hypothyroidism, hydrocephalus, seizure disorders such as Sydenham's chorea and cerebral palsy, spina bifida, anencephaly, craniorachischisis, congenital glaucoma, cataract, age-related macular degeneration, and sensorineural hearing loss; an epithelial disorder, such as dyshidrotic eczema, allergic contact dermatitis, keratosis pilaris, melasma, vitiligo, actinic keratosis, basal cell carcinoma, squamous cell carcinoma, seborrheic keratosis, folliculitis, herpes simplex, herpes zoster, varicella, candidiasis, dermatophytosis, scabies, insect bites, cherry angioma, keloid, dermatofibroma, acrochordons, urticaria, transient acantholytic dermatosis, xerosis, eczema, atopic dermatitis, contact dermatitis, hand eczema, nummular eczema, lichen simplex chronicus, asteatotic eczema, stasis dermatitis and stasis ulceration, seborrheic dermatitis, psoriasis, lichen planus, pityriasis rosea, impetigo, ecthyma, dermatophytosis, tinea versicolor, warts, acne vulgaris, acne rosacea, pemphigus vulgaris, pemphigus foliaceus, paraneoplastic pemphigus, bullous pemphigoid, herpes gestationis, dermatitis herpetiformis, linear IgA disease, epidermolysis bullosa acquisita, dermatomyositis, lupus erythematosus, scleroderma and morphea, erythroderma, alopecia, figurate skin lesions, telangiectasias, hypopigmentation, hyperpigmentation, vesicles/bullae, exanthems, cutaneous drug reactions, papulonodular skin lesions, chronic non-healing wounds, photosensitivity diseases, epidermolysis bullosa simplex, epidermolytic hyperkeratosis, epidermolytic and nonepidermolytic palmoplantar keratoderma, ichthyosis bullosa of Siemens, ichthyosis exfoliativa, keratosis palmaris et plantaris, keratosis palmoplantar, palmoplantar keratoderma, keratosis punctata, Meesmann's corneal dystrophy, pachyonychia congenita, white sponge nevus, steatocystoma multiplex, epidermal nevi/epidermolytic hyperkeratosis type, monilethrix, trichothiodystrophy, chronic hepatitis/cryptogenic cirrhosis, and colorectal hyperplasia; a neurological disorder, such as epilepsy, ischemic cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's disease, Pick's disease, Huntington's disease, dementia, Parkinson's disease and other extrapyramidal disorders, amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease, prion

diseases including kuru, Creutzfeldt-Jakob disease, and Gerstmann-Straussler-Scheinker syndrome, fatal familial insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis, encephalotrigeminal syndrome, mental retardation and other developmental disorders of the central nervous system including Down syndrome,

5 cerebral palsy, neuroskeletal disorders, autonomic nervous system disorders, cranial nerve disorders, spinal cord diseases, muscular dystrophy and other neuromuscular disorders, peripheral nervous system disorders, dermatomyositis and polymyositis, inherited, metabolic, endocrine, and toxic myopathies, myasthenia gravis, periodic paralysis, mental disorders including mood, anxiety, and schizophrenic disorders, seasonal affective disorder (SAD), akathesia, amnesia, catatonia, diabetic neuropathy, tardive

10 dyskinesia, dystonias, paranoid psychoses, postherpetic neuralgia, Tourette's disorder, progressive supranuclear palsy, corticobasal degeneration, and familial frontotemporal dementia; and a reproductive disorder, such as infertility, including tubal disease, ovulatory defects, and endometriosis, a disorder of prolactin production, a disruption of the estrous cycle, a disruption of the menstrual cycle, polycystic ovary syndrome, ovarian hyperstimulation syndrome, an endometrial or ovarian tumor, a uterine

15 fibroid, autoimmune disorders, an ectopic pregnancy, and teratogenesis; cancer of the breast, fibrocystic breast disease, and galactorrhea; a disruption of spermatogenesis, abnormal sperm physiology, cancer of the testis, cancer of the prostate, benign prostatic hyperplasia, prostatitis, Peyronie's disease, impotence, carcinoma of the male breast, and gynecomastia. The polynucleotide sequences encoding PRTS may be used in Southern or northern analysis, dot blot, or other membrane-based technologies; in

20 PCR technologies; in dipstick, pin, and multiformat ELISA-like assays; and in microarrays utilizing fluids or tissues from patients to detect altered PRTS expression. Such qualitative or quantitative methods are well-known in the art.

In a particular aspect, the nucleotide sequences encoding PRTS may be useful in assays that detect the presence of associated disorders, particularly those mentioned above. The nucleotide

25 sequences encoding PRTS may be labeled by standard methods and added to a fluid or tissue sample from a patient under conditions suitable for the formation of hybridization complexes. After a suitable incubation period, the sample is washed and the signal is quantified and compared with a standard value. If the amount of signal in the patient sample is significantly altered in comparison to a control sample then the presence of altered levels of nucleotide sequences encoding PRTS in the sample

30 indicates the presence of the associated disorder. Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies, in clinical trials, or to monitor the treatment of an individual patient.

In order to provide a basis for the diagnosis of a disorder associated with expression of PRTS, a normal or standard profile for expression is established. This may be accomplished by combining

body fluids or cell extracts taken from normal subjects, either animal or human, with a sequence, or a fragment thereof, encoding PRTS, under conditions suitable for hybridization or amplification.

Standard hybridization may be quantified by comparing the values obtained from normal subjects with values from an experiment in which a known amount of a substantially purified polynucleotide is used.

- 5 Standard values obtained in this manner may be compared with values obtained from samples from patients who are symptomatic for a disorder. Deviation from standard values is used to establish the presence of a disorder.

Once the presence of a disorder is established and a treatment protocol is initiated, hybridization assays may be repeated on a regular basis to determine if the level of expression in the patient begins to approximate that which is observed in the normal subject. The results obtained from successive assays may be used to show the efficacy of treatment over a period ranging from several days to months.

With respect to cancer, the presence of an abnormal amount of transcript (either under- or overexpressed) in biopsied tissue from an individual may indicate a predisposition for the development of the disease, or may provide a means for detecting the disease prior to the appearance of actual clinical symptoms. A more definitive diagnosis of this type may allow health professionals to employ preventative measures or aggressive treatment earlier thereby preventing the development or further progression of the cancer.

Additional diagnostic uses for oligonucleotides designed from the sequences encoding PRTS 20 may involve the use of PCR. These oligomers may be chemically synthesized, generated enzymatically, or produced in vitro. Oligomers will preferably contain a fragment of a polynucleotide encoding PRTS, or a fragment of a polynucleotide complementary to the polynucleotide encoding PRTS, and will be employed under optimized conditions for identification of a specific gene or condition. Oligomers may also be employed under less stringent conditions for detection or quantification of closely related DNA 25 or RNA sequences.

In a particular aspect, oligonucleotide primers derived from the polynucleotide sequences encoding PRTS may be used to detect single nucleotide polymorphisms (SNPs). SNPs are substitutions, insertions and deletions that are a frequent cause of inherited or acquired genetic disease in humans. Methods of SNP detection include, but are not limited to, single-stranded conformation 30 polymorphism (SSCP) and fluorescent SSCP (fSSCP) methods. In SSCP, oligonucleotide primers derived from the polynucleotide sequences encoding PRTS are used to amplify DNA using the polymerase chain reaction (PCR). The DNA may be derived, for example, from diseased or normal tissue, biopsy samples, bodily fluids, and the like. SNPs in the DNA cause differences in the secondary and tertiary structures of PCR products in single-stranded form, and these differences are detectable

using gel electrophoresis in non-denaturing gels. In fSCCP, the oligonucleotide primers are fluorescently labeled, which allows detection of the amplimers in high-throughput equipment such as DNA sequencing machines. Additionally, sequence database analysis methods, termed *in silico* SNP (isSNP), are capable of identifying polymorphisms by comparing the sequence of individual

5 overlapping DNA fragments which assemble into a common consensus sequence. These computer-based methods filter out sequence variations due to laboratory preparation of DNA and sequencing errors using statistical models and automated analyses of DNA sequence chromatograms. In the alternative, SNPs may be detected and characterized by mass spectrometry using, for example, the high throughput MASSARRAY system (Sequenom, Inc., San Diego CA).

10 Methods which may also be used to quantify the expression of PRTS include radiolabeling or biotinyling nucleotides, coamplification of a control nucleic acid, and interpolating results from standard curves. (See, e.g., Melby, P.C. et al. (1993) J. Immunol. Methods 159:235-244; Duplaa, C. et al. (1993) Anal. Biochem. 212:229-236.) The speed of quantitation of multiple samples may be accelerated by running the assay in a high-throughput format where the oligomer or polynucleotide of

15 interest is presented in various dilutions and a spectrophotometric or colorimetric response gives rapid quantitation.

In further embodiments, oligonucleotides or longer fragments derived from any of the polynucleotide sequences described herein may be used as elements on a microarray. The microarray can be used in transcript imaging techniques which monitor the relative expression levels of large numbers of genes simultaneously as described below. The microarray may also be used to identify genetic variants, mutations, and polymorphisms. This information may be used to determine gene function, to understand the genetic basis of a disorder, to diagnose a disorder, to monitor progression/regression of disease as a function of gene expression, and to develop and monitor the activities of therapeutic agents in the treatment of disease. In particular, this information may be used to develop a pharmacogenomic profile of a patient in order to select the most appropriate and effective treatment regimen for that patient. For example, therapeutic agents which are highly effective and display the fewest side effects may be selected for a patient based on his/her pharmacogenomic profile.

In another embodiment, PRTS, fragments of PRTS, or antibodies specific for PRTS may be used as elements on a microarray. The microarray may be used to monitor or measure protein-protein interactions, drug-target interactions, and gene expression profiles, as described above.

A particular embodiment relates to the use of the polynucleotides of the present invention to generate a transcript image of a tissue or cell type. A transcript image represents the global pattern of gene expression by a particular tissue or cell type. Global gene expression patterns are analyzed by quantifying the number of expressed genes and their relative abundance under given conditions and at a

given time. (See Seilhamer et al., "Comparative Gene Transcript Analysis," U.S. Patent Number 5,840,484, expressly incorporated by reference herein.) Thus a transcript image may be generated by hybridizing the polynucleotides of the present invention or their complements to the totality of transcripts or reverse transcripts of a particular tissue or cell type. In one embodiment, the 5 hybridization takes place in high-throughput format, wherein the polynucleotides of the present invention or their complements comprise a subset of a plurality of elements on a microarray. The resultant transcript image would provide a profile of gene activity.

Transcript images may be generated using transcripts isolated from tissues, cell lines, biopsies, or other biological samples. The transcript image may thus reflect gene expression in vivo, as in the 10 case of a tissue or biopsy sample, or in vitro, as in the case of a cell line.

Transcript images which profile the expression of the polynucleotides of the present invention may also be used in conjunction with in vitro model systems and preclinical evaluation of pharmaceuticals, as well as toxicological testing of industrial and naturally-occurring environmental compounds. All compounds induce characteristic gene expression patterns, frequently termed 15 molecular fingerprints or toxicant signatures, which are indicative of mechanisms of action and toxicity (Nuwaysir, E.F. et al. (1999) Mol. Carcinog. 24:153-159; Steiner, S. and N.L. Anderson (2000) Toxicol. Lett. 112-113:467-471, expressly incorporated by reference herein). If a test compound has a signature similar to that of a compound with known toxicity, it is likely to share those toxic properties. These fingerprints or signatures are most useful and refined when they contain expression information 20 from a large number of genes and gene families. Ideally, a genome-wide measurement of expression provides the highest quality signature. Even genes whose expression is not altered by any tested compounds are important as well, as the levels of expression of these genes are used to normalize the rest of the expression data. The normalization procedure is useful for comparison of expression data after treatment with different compounds. While the assignment of gene function to elements of a 25 toxicant signature aids in interpretation of toxicity mechanisms, knowledge of gene function is not necessary for the statistical matching of signatures which leads to prediction of toxicity. (See, for example, Press Release 00-02 from the National Institute of Environmental Health Sciences, released February 29, 2000, available at <http://www.niehs.nih.gov/oc/news/toxchip.htm>.) Therefore, it is important and desirable in toxicological screening using toxicant signatures to include all expressed 30 gene sequences.

In one embodiment, the toxicity of a test compound is assessed by treating a biological sample containing nucleic acids with the test compound. Nucleic acids that are expressed in the treated biological sample are hybridized with one or more probes specific to the polynucleotides of the present invention, so that transcript levels corresponding to the polynucleotides of the present

invention may be quantified. The transcript levels in the treated biological sample are compared with levels in an untreated biological sample. Differences in the transcript levels between the two samples are indicative of a toxic response caused by the test compound in the treated sample.

Another particular embodiment relates to the use of the polypeptide sequences of the present invention to analyze the proteome of a tissue or cell type. The term proteome refers to the global pattern of protein expression in a particular tissue or cell type. Each protein component of a proteome can be subjected individually to further analysis. Proteome expression patterns, or profiles, are analyzed by quantifying the number of expressed proteins and their relative abundance under given conditions and at a given time. A profile of a cell's proteome may thus be generated by separating and analyzing the polypeptides of a particular tissue or cell type. In one embodiment, the separation is achieved using two-dimensional gel electrophoresis, in which proteins from a sample are separated by isoelectric focusing in the first dimension, and then according to molecular weight by sodium dodecyl sulfate slab gel electrophoresis in the second dimension (Steiner and Anderson, *supra*). The proteins are visualized in the gel as discrete and uniquely positioned spots, typically by staining the gel with an agent such as Coomassie Blue or silver or fluorescent stains. The optical density of each protein spot is generally proportional to the level of the protein in the sample. The optical densities of equivalently positioned protein spots from different samples, for example, from biological samples either treated or untreated with a test compound or therapeutic agent, are compared to identify any changes in protein spot density related to the treatment. The proteins in the spots are partially sequenced using, for example, standard methods employing chemical or enzymatic cleavage followed by mass spectrometry. The identity of the protein in a spot may be determined by comparing its partial sequence, preferably of at least 5 contiguous amino acid residues, to the polypeptide sequences of the present invention. In some cases, further sequence data may be obtained for definitive protein identification.

A proteomic profile may also be generated using antibodies specific for PRTS to quantify the levels of PRTS expression. In one embodiment, the antibodies are used as elements on a microarray, and protein expression levels are quantified by exposing the microarray to the sample and detecting the levels of protein bound to each array element (Lueking, A. et al. (1999) *Anal. Biochem.* 270:103-111; Mendoza, L.G. et al. (1999) *Biotechniques* 27:778-788). Detection may be performed by a variety of methods known in the art, for example, by reacting the proteins in the sample with a thiol- or amino-reactive fluorescent compound and detecting the amount of fluorescence bound at each array element.

Toxicant signatures at the proteome level are also useful for toxicological screening, and should be analyzed in parallel with toxicant signatures at the transcript level. There is a poor correlation between transcript and protein abundances for some proteins in some tissues (Anderson, N.L. and J. Seilhamer (1997) *Electrophoresis* 18:533-537), so proteome toxicant signatures may be useful in the

analysis of compounds which do not significantly affect the transcript image, but which alter the proteomic profile. In addition, the analysis of transcripts in body fluids is difficult, due to rapid degradation of mRNA, so proteomic profiling may be more reliable and informative in such cases.

In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins that are expressed in the treated biological sample are separated so that the amount of each protein can be quantified. The amount of each protein is compared to the amount of the corresponding protein in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample. Individual proteins are identified by sequencing the amino acid residues of the individual proteins and comparing these partial sequences to the polypeptides of the present invention.

In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins from the biological sample are incubated with antibodies specific to the polypeptides of the present invention. The amount of protein recognized by the antibodies is quantified. The amount of protein in the treated biological sample is compared with the amount in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample.

Microarrays may be prepared, used, and analyzed using methods known in the art. (See, e.g., Brennan, T.M. et al. (1995) U.S. Patent No. 5,474,796; Schena, M. et al. (1996) Proc. Natl. Acad. Sci. USA 93:10614-10619; Baldeschweiler et al. (1995) PCT application WO95/251116; Shalon, D. et al. (1995) PCT application WO95/35505; Heller, R.A. et al. (1997) Proc. Natl. Acad. Sci. USA 94:2150-2155; and Heller, M.J. et al. (1997) U.S. Patent No. 5,605,662.) Various types of microarrays are well known and thoroughly described in DNA Microarrays: A Practical Approach, M. Schena, ed. (1999) Oxford University Press, London, hereby expressly incorporated by reference.

In another embodiment of the invention, nucleic acid sequences encoding PRTS may be used to generate hybridization probes useful in mapping the naturally occurring genomic sequence. Either coding or noncoding sequences may be used, and in some instances, noncoding sequences may be preferable over coding sequences. For example, conservation of a coding sequence among members of a multi-gene family may potentially cause undesired cross hybridization during chromosomal mapping. The sequences may be mapped to a particular chromosome, to a specific region of a chromosome, or to artificial chromosome constructions, e.g., human artificial chromosomes (HACs), yeast artificial chromosomes (YACs), bacterial artificial chromosomes (BACs), bacterial P1 constructions, or single chromosome cDNA libraries. (See, e.g., Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355; Price, C.M. (1993) Blood Rev. 7:127-134; and Trask, B.J. (1991) Trends Genet. 7:149-154.) Once mapped, the nucleic acid sequences of the invention may be used to develop genetic

linkage maps, for example, which correlate the inheritance of a disease state with the inheritance of a particular chromosome region or restriction fragment length polymorphism (RFLP). (See, for example, Lander, E.S. and D. Botstein (1986) Proc. Natl. Acad. Sci. USA 83:7353-7357.)

Fluorescent in situ hybridization (FISH) may be correlated with other physical and genetic map data. (See, e.g., Heinz-Ulrich, et al. (1995) in Meyers, supra, pp. 965-968.) Examples of genetic map data can be found in various scientific journals or at the Online Mendelian Inheritance in Man (OMIM) World Wide Web site. Correlation between the location of the gene encoding PRTS on a physical map and a specific disorder, or a predisposition to a specific disorder, may help define the region of DNA associated with that disorder and thus may further positional cloning efforts.

In situ hybridization of chromosomal preparations and physical mapping techniques, such as linkage analysis using established chromosomal markers, may be used for extending genetic maps. Often the placement of a gene on the chromosome of another mammalian species, such as mouse, may reveal associated markers even if the exact chromosomal locus is not known. This information is valuable to investigators searching for disease genes using positional cloning or other gene discovery techniques. Once the gene or genes responsible for a disease or syndrome have been crudely localized by genetic linkage to a particular genomic region, e.g., ataxia-telangiectasia to 11q22-23, any sequences mapping to that area may represent associated or regulatory genes for further investigation. (See, e.g., Gatti, R.A. et al. (1988) Nature 336:577-580.) The nucleotide sequence of the instant invention may also be used to detect differences in the chromosomal location due to translocation, inversion, etc., among normal, carrier, or affected individuals.

In another embodiment of the invention, PRTS, its catalytic or immunogenic fragments, or oligopeptides thereof can be used for screening libraries of compounds in any of a variety of drug screening techniques. The fragment employed in such screening may be free in solution, affixed to a solid support, borne on a cell surface, or located intracellularly. The formation of binding complexes between PRTS and the agent being tested may be measured.

Another technique for drug screening provides for high throughput screening of compounds having suitable binding affinity to the protein of interest. (See, e.g., Geysen, et al. (1984) PCT application WO84/03564.) In this method, large numbers of different small test compounds are synthesized on a solid substrate. The test compounds are reacted with PRTS, or fragments thereof, and washed. Bound PRTS is then detected by methods well known in the art. Purified PRTS can also be coated directly onto plates for use in the aforementioned drug screening techniques. Alternatively, non-neutralizing antibodies can be used to capture the peptide and immobilize it on a solid support.

In another embodiment, one may use competitive drug screening assays in which neutralizing antibodies capable of binding PRTS specifically compete with a test compound for binding PRTS. In

this manner, antibodies can be used to detect the presence of any peptide which shares one or more antigenic determinants with PRTS.

In additional embodiments, the nucleotide sequences which encode PRTS may be used in any molecular biology techniques that have yet to be developed, provided the new techniques rely on

5 properties of nucleotide sequences that are currently known, including, but not limited to, such properties as the triplet genetic code and specific base pair interactions.

Without further elaboration, it is believed that one skilled in the art can, using the preceding description, utilize the present invention to its fullest extent. The following embodiments are, therefore, to be construed as merely illustrative, and not limitative of the remainder of the disclosure

10 in any way whatsoever.

The disclosures of all patents, applications, and publications mentioned above and below, including U.S. Ser. No. 60/212,336, U.S. Ser. No. 60/213,995, U.S. Ser. No. 60/215,396, U.S. Ser. No. 60/216,821, and U.S. Ser. No. 60/218,946, are hereby expressly incorporated by reference.

15

## EXAMPLES

### I. Construction of cDNA Libraries

Incyte cDNAs were derived from cDNA libraries described in the LIFESEQ GOLD database (Incyte Genomics, Palo Alto CA) and shown in Table 4, column 5. Some tissues were homogenized and lysed in guanidinium isothiocyanate, while others were homogenized and lysed in phenol or in a

20 suitable mixture of denaturants, such as TRIZOL (Life Technologies), a monophasic solution of phenol and guanidine isothiocyanate. The resulting lysates were centrifuged over CsCl cushions or extracted with chloroform. RNA was precipitated from the lysates with either isopropanol or sodium acetate and ethanol, or by other routine methods.

Phenol extraction and precipitation of RNA were repeated as necessary to increase RNA

25 purity. In some cases, RNA was treated with DNase. For most libraries, poly(A)+ RNA was isolated using oligo d(T)-coupled paramagnetic particles (Promega), OLIGOTEX latex particles (QIAGEN, Chatsworth CA), or an OLIGOTEX mRNA purification kit (QIAGEN). Alternatively, RNA was isolated directly from tissue lysates using other RNA isolation kits, e.g., the POLY(A)PURE mRNA purification kit (Ambion, Austin TX).

30 In some cases, Stratagene was provided with RNA and constructed the corresponding cDNA libraries. Otherwise, cDNA was synthesized and cDNA libraries were constructed with the UNIZAP vector system (Stratagene) or SUPERSCRIPT plasmid system (Life Technologies), using the recommended procedures or similar methods known in the art. (See, e.g., Ausubel, 1997, supra, units 5.1-6.6.) Reverse transcription was initiated using oligo d(T) or random primers. Synthetic

oligonucleotide adapters were ligated to double stranded cDNA, and the cDNA was digested with the appropriate restriction enzyme or enzymes. For most libraries, the cDNA was size-selected (300-1000 bp) using SEPHACRYL S1000, SEPHAROSE CL2B, or SEPHAROSE CL4B column chromatography (Amersham Pharmacia Biotech) or preparative agarose gel electrophoresis. cDNAs 5 were ligated into compatible restriction enzyme sites of the polylinker of a suitable plasmid, e.g., PBLUESCRIPT plasmid (Stratagene), PSSPORT1 plasmid (Life Technologies), PCDNA2.1 plasmid (Invitrogen, Carlsbad CA), PBK-CMV plasmid (Stratagene), or pINCY (Incyte Genomics, Palo Alto CA), or derivatives thereof. Recombinant plasmids were transformed into competent *E. coli* cells including XL1-Blue, XL1-BlueMRF, or SOLR from Stratagene or DH5 $\alpha$ , DH10B, or ElectroMAX 10 DH10B from Life Technologies.

## II. Isolation of cDNA Clones

Plasmids obtained as described in Example I were recovered from host cells by *in vivo* excision using the UNIZAP vector system (Stratagene) or by cell lysis. Plasmids were purified using at least one of the following: a Magic or WIZARD Minipreps DNA purification system (Promega); an AGTC 15 Miniprep purification kit (Edge Biosystems, Gaithersburg MD); and QIAWELL 8 Plasmid, QIAWELL 8-Plus Plasmid, QIAWELL 8-Ultra Plasmid purification systems or the R.E.A.L. PREP 96 plasmid purification kit from QIAGEN. Following precipitation, plasmids were resuspended in 0.1 ml of distilled water and stored, with or without lyophilization, at 4 °C.

Alternatively, plasmid DNA was amplified from host cell lysates using direct link PCR in a 20 high-throughput format (Rao, V.B.: (1994) Anal. Biochem. 216:1-14). Host cell lysis and thermal cycling steps were carried out in a single reaction mixture. Samples were processed and stored in 384-well plates, and the concentration of amplified plasmid DNA was quantified fluorometrically using 25 PICOGREEN dye (Molecular Probes, Eugene OR) and a FLUOROSCAN II fluorescence scanner (Labsystems Oy, Helsinki, Finland).

## III. Sequencing and Analysis

Incyte cDNA recovered in plasmids as described in Example II were sequenced as follows. Sequencing reactions were processed using standard methods or high-throughput instrumentation such as the ABI CATALYST 800 (Applied Biosystems) thermal cycler or the PTC-200 thermal cycler (MJ Research) in conjunction with the HYDRA microdispenser (Robbins Scientific) or the 30 MICROLAB 2200 (Hamilton) liquid transfer system. cDNA sequencing reactions were prepared using reagents provided by Amersham Pharmacia Biotech or supplied in ABI sequencing kits such as the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems). Electrophoretic separation of cDNA sequencing reactions and detection of labeled polynucleotides were carried out using the MEGABACE 1000 DNA sequencing system (Molecular Dynamics); the ABI

PRISM 373 or 377 sequencing system (Applied Biosystems) in conjunction with standard ABI protocols and base calling software; or other sequence analysis systems known in the art. Reading frames within the cDNA sequences were identified using standard methods (reviewed in Ausubel, 1997, supra, unit 7.7). Some of the cDNA sequences were selected for extension using the techniques disclosed in Example VIII.

The polynucleotide sequences derived from Incyte cDNAs were validated by removing vector, linker, and poly(A) sequences and by masking ambiguous bases, using algorithms and programs based on BLAST, dynamic programming, and dinucleotide nearest neighbor analysis. The Incyte cDNA sequences or translations thereof were then queried against a selection of public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases, and BLOCKS, PRINTS, DOMO, PRODOM, and hidden Markov model (HMM)-based protein family databases such as PFAM. (HMM is a probabilistic approach which analyzes consensus primary structures of gene families. See, for example, Eddy, S.R. (1996) Curr. Opin. Struct. Biol. 6:361-365.) The queries were performed using programs based on BLAST, FASTA, BLIMPS, and HMMER. The Incyte cDNA sequences were assembled to produce full length polynucleotide sequences. Alternatively, GenBank cDNAs, GenBank ESTs, stitched sequences, stretched sequences, or Genscan-predicted coding sequences (see Examples IV and V) were used to extend Incyte cDNA assemblages to full length. Assembly was performed using programs based on Phred, Phrap, and Consed, and cDNA assemblages were screened for open reading frames using programs based on GeneMark, BLAST, and FASTA.

The full length polynucleotide sequences were translated to derive the corresponding full length polypeptide sequences. Alternatively, a polypeptide of the invention may begin at any of the methionine residues of the full length translated polypeptide. Full length polypeptide sequences were subsequently analyzed by querying against databases such as the GenBank protein databases (genpept), SwissProt, BLOCKS, PRINTS, DOMO, PRODOM, Prosite, and hidden Markov model (HMM)-based protein family databases such as PFAM. Full length polynucleotide sequences are also analyzed using MACDNASIS PRO software (Hitachi Software Engineering, South San Francisco CA) and LASERGENE software (DNASTAR). Polynucleotide and polypeptide sequence alignments are generated using default parameters specified by the CLUSTAL algorithm as incorporated into the MEGALIGN multisequence alignment program (DNASTAR), which also calculates the percent identity between aligned sequences.

Table 7 summarizes the tools, programs, and algorithms used for the analysis and assembly of Incyte cDNA and full length sequences and provides applicable descriptions, references, and threshold parameters. The first column of Table 7 shows the tools, programs, and algorithms used, the second column provides brief descriptions thereof, the third column presents appropriate references, all of

which are incorporated by reference herein in their entirety, and the fourth column presents, where applicable, the scores, probability values, and other parameters used to evaluate the strength of a match between two sequences (the higher the score or the lower the probability value, the greater the identity between two sequences).

5       The programs described above for the assembly and analysis of full length polynucleotide and polypeptide sequences were also used to identify polynucleotide sequence fragments from SEQ ID NO:22-42. Fragments from about 20 to about 4000 nucleotides which are useful in hybridization and amplification technologies are described in Table 4, column 4.

#### IV. Identification and Editing of Coding Sequences from Genomic DNA

10     Putative proteases were initially identified by running the Genscan gene identification program against public genomic sequence databases (e.g., gbpri and gbhtg). Genscan is a general-purpose gene identification program which analyzes genomic DNA sequences from a variety of organisms (See Burge, C. and S. Karlin (1997) J. Mol. Biol. 268:78-94, and Burge, C. and S. Karlin (1998) Curr. Opin. Struct. Biol. 8:346-354). The program concatenates predicted exons to form an assembled  
15     cDNA sequence extending from a methionine to a stop codon. The output of Genscan is a FASTA database of polynucleotide and polypeptide sequences. The maximum range of sequence for Genscan to analyze at once was set to 30 kb. To determine which of these Genscan predicted cDNA sequences encode proteases, the encoded polypeptides were analyzed by querying against PFAM models for proteases. Potential proteases were also identified by homology to Incyte cDNA sequences that had  
20     been annotated as proteases. These selected Genscan-predicted sequences were then compared by BLAST analysis to the genpept and gbpri public databases. Where necessary, the Genscan-predicted sequences were then edited by comparison to the top BLAST hit from genpept to correct errors in the sequence predicted by Genscan, such as extra or omitted exons. BLAST analysis was also used to find any Incyte cDNA or public cDNA coverage of the Genscan-predicted sequences, thus providing  
25     evidence for transcription. When Incyte cDNA coverage was available, this information was used to correct or confirm the Genscan predicted sequence. Full length polynucleotide sequences were obtained by assembling Genscan-predicted coding sequences with Incyte cDNA sequences and/or public cDNA sequences using the assembly process described in Example III. Alternatively, full length polynucleotide sequences were derived entirely from edited or unedited Genscan-predicted coding  
30     sequences.

#### V. Assembly of Genomic Sequence Data with cDNA Sequence Data

##### "Stitched" Sequences

Partial cDNA sequences were extended with exons predicted by the Genscan gene identification program described in Example IV. Partial cDNAs assembled as described in Example III were mapped

to genomic DNA and parsed into clusters containing related cDNAs and Genscan exon predictions from one or more genomic sequences. Each cluster was analyzed using an algorithm based on graph theory and dynamic programming to integrate cDNA and genomic information, generating possible splice variants that were subsequently confirmed, edited, or extended to create a full length sequence.

5 Sequence intervals in which the entire length of the interval was present on more than one sequence in the cluster were identified, and intervals thus identified were considered to be equivalent by transitivity. For example, if an interval was present on a cDNA and two genomic sequences, then all three intervals were considered to be equivalent. This process allows unrelated but consecutive genomic sequences to be brought together, bridged by cDNA sequence. Intervals thus identified were then "stitched" together

10 by the stitching algorithm in the order that they appear along their parent sequences to generate the longest possible sequence, as well as sequence variants. Linkages between intervals which proceed along one type of parent sequence (cDNA to cDNA or genomic sequence to genomic sequence) were given preference over linkages which change parent type (cDNA to genomic sequence). The resultant stitched sequences were translated and compared by BLAST analysis to the genpept and gbpri public

15 databases. Incorrect exons predicted by Genscan were corrected by comparison to the top BLAST hit from genpept. Sequences were further extended with additional cDNA sequences, or by inspection of genomic DNA, when necessary.

#### "Stretched" Sequences

Partial DNA sequences were extended to full length with an algorithm based on BLAST

20 analysis. First, partial cDNAs assembled as described in Example III were queried against public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases using the BLAST program. The nearest GenBank protein homolog was then compared by BLAST analysis to either Incyte cDNA sequences or GenScan exon predicted sequences described in Example IV. A chimeric protein was generated by using the resultant high-scoring segment pairs (HSPs) to map the

25 translated sequences onto the GenBank protein homolog. Insertions or deletions may occur in the chimeric protein with respect to the original GenBank protein homolog. The GenBank protein homolog, the chimeric protein, or both were used as probes to search for homologous genomic sequences from the public human genome databases. Partial DNA sequences were therefore "stretched" or extended by the addition of homologous genomic sequences. The resultant stretched sequences were examined to

30 determine whether it contained a complete gene.

#### VI. Chromosomal Mapping of PRTS Encoding Polynucleotides

The sequences which were used to assemble SEQ ID NO:22-42 were compared with sequences from the Incyte LIFESEQ database and public domain databases using BLAST and other implementations of the Smith-Waterman algorithm. Sequences from these databases that matched

SEQ ID NO:22-42 were assembled into clusters of contiguous and overlapping sequences using assembly algorithms such as Phrap (Table 7). Radiation hybrid and genetic mapping data available from public resources such as the Stanford Human Genome Center (SHGC), Whitehead Institute for Genome Research (WIGR), and Généthon were used to determine if any of the clustered sequences 5 had been previously mapped. Inclusion of a mapped sequence in a cluster resulted in the assignment of all sequences of that cluster, including its particular SEQ ID NO:, to that map location.

Map locations are represented by ranges, or intervals, of human chromosomes. The map position of an interval, in centiMorgans, is measured relative to the terminus of the chromosome's p-arm. (The centiMorgan (cM) is a unit of measurement based on recombination frequencies between 10 chromosomal markers. On average, 1 cM is roughly equivalent to 1 megabase (Mb) of DNA in humans, although this can vary widely due to hot and cold spots of recombination.) The cM distances are based on genetic markers mapped by Généthon which provide boundaries for radiation hybrid markers whose sequences were included in each of the clusters. Human genome maps and other resources available to the public, such as the NCBI "GeneMap'99" World Wide Web site 15 (<http://www.ncbi.nlm.nih.gov/genemap/>), can be employed to determine if previously identified disease genes map within or in proximity to the intervals indicated above.

In this manner, SEQ ID NO:25 was mapped to chromosome 5 within the interval from 69.60 to 76.50 centiMorgans. SEQ ID NO:28 was mapped to chromosome 16 within the interval from 16 81.80 to 84.40 centiMorgans.

## 20 VII. Analysis of Polynucleotide Expression

Northern analysis is a laboratory technique used to detect the presence of a transcript of a gene and involves the hybridization of a labeled nucleotide sequence to a membrane on which RNAs from a particular cell type or tissue have been bound. (See, e.g., Sambrook, supra, ch. 7; Ausubel (1995) supra, ch. 4 and 16.)

25 Analogous computer techniques applying BLAST were used to search for identical or related molecules in cDNA databases such as GenBank or LIFESEQ (Incyte Genomics). This analysis is much faster than multiple membrane-based hybridizations. In addition, the sensitivity of the computer search can be modified to determine whether any particular match is categorized as exact or similar. The basis of the search is the product score, which is defined as:

30

$$\frac{\text{BLAST Score} \times \text{Percent Identity}}{5 \times \min\{\text{length(Seq. 1)}, \text{length(Seq. 2)}\}}$$

The product score takes into account both the degree of similarity between two sequences and the length 35 of the sequence match. The product score is a normalized value between 0 and 100, and is calculated

as follows: the BLAST score is multiplied by the percent nucleotide identity and the product is divided by (5 times the length of the shorter of the two sequences). The BLAST score is calculated by assigning a score of +5 for every base that matches in a high-scoring segment pair (HSP), and -4 for every mismatch. Two sequences may share more than one HSP (separated by gaps). If there is more  
5 than one HSP, then the pair with the highest BLAST score is used to calculate the product score. The product score represents a balance between fractional overlap and quality in a BLAST alignment. For example, a product score of 100 is produced only for 100% identity over the entire length of the shorter of the two sequences being compared. A product score of 70 is produced either by 100% identity and 70% overlap at one end, or by 88% identity and 100% overlap at the other. A product score of 50 is  
10 produced either by 100% identity and 50% overlap at one end, or 79% identity and 100% overlap.

Alternatively, polynucleotide sequences encoding PRTS are analyzed with respect to the tissue sources from which they were derived. For example, some full length sequences are assembled, at least in part, with overlapping Incyte cDNA sequences (see Example III). Each cDNA sequence is derived from a cDNA library constructed from a human tissue. Each human tissue is classified into one of the  
15 following organ/tissue categories: cardiovascular system; connective tissue; digestive system; embryonic structures; endocrine system; exocrine glands; genitalia, female; genitalia, male; germ cells; hemic and immune system; liver; musculoskeletal system; nervous system; pancreas; respiratory system; sense organs; skin; stomatognathic system; unclassified/mixed; or urinary tract. The number of libraries in each category is counted and divided by the total number of libraries across all categories.  
20 Similarly, each human tissue is classified into one of the following disease/condition categories: cancer, cell line, developmental, inflammation, neurological, trauma, cardiovascular, pooled, and other, and the number of libraries in each category is counted and divided by the total number of libraries across all categories. The resulting percentages reflect the tissue- and disease-specific expression of cDNA encoding PRTS. cDNA sequences and cDNA library/tissue information are found in the LIFESEQ  
25 GOLD database (Incyte Genomics, Palo Alto CA).

#### VIII. Extension of PRTS Encoding Polynucleotides

Full length polynucleotide sequences were also produced by extension of an appropriate fragment of the full length molecule using oligonucleotide primers designed from this fragment. One primer was synthesized to initiate 5' extension of the known fragment, and the other primer was  
30 synthesized to initiate 3' extension of the known fragment. The initial primers were designed using OLIGO 4.06 software (National Biosciences), or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the target sequence at temperatures of about 68 °C to about 72 °C. Any stretch of nucleotides which would result in hairpin structures and primer-primer dimerizations was avoided.

Selected human cDNA libraries were used to extend the sequence. If more than one extension was necessary or desired, additional or nested sets of primers were designed.

High fidelity amplification was obtained by PCR using methods well known in the art. PCR was performed in 96-well plates using the PTC-200 thermal cycler (MJ Research, Inc.). The reaction mix contained DNA template, 200 nmol of each primer, reaction buffer containing Mg<sup>2+</sup>, (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, and 2-mercaptoethanol, Taq DNA polymerase (Amersham Pharmacia Biotech), ELONGASE enzyme (Life Technologies), and Pfu DNA polymerase (Stratagene), with the following parameters for primer pair PCI A and PCI B: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C. In the alternative, the parameters for primer pair T7 and SK+ were as follows: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 57°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C.

The concentration of DNA in each well was determined by dispensing 100 µl PICOGREEN quantitation reagent (0.25% (v/v) PICOGREEN; Molecular Probes, Eugene OR) dissolved in 1X TE and 0.5 µl of undiluted PCR product into each well of an opaque fluorimeter plate (Corning Costar, Acton MA); allowing the DNA to bind to the reagent. The plate was scanned in a Fluoroskan II (Labsystems Oy, Helsinki, Finland) to measure the fluorescence of the sample and to quantify the concentration of DNA. A 5 µl to 10 µl aliquot of the reaction mixture was analyzed by electrophoresis on a 1% agarose gel to determine which reactions were successful in extending the sequence.

20 The extended nucleotides were desalted and concentrated, transferred to 384-well plates, digested with C<sub>vi</sub>JI cholera virus endonuclease (Molecular Biology Research, Madison WI), and sonicated or sheared prior to religation into pUC 18 vector (Amersham Pharmacia Biotech). For shotgun sequencing, the digested nucleotides were separated on low concentration (0.6 to 0.8%) agarose gels; fragments were excised, and agar digested with Agar ACE (Promega). Extended clones were 25 religated using T4 ligase (New England Biolabs, Beverly MA) into pUC 18 vector (Amersham Pharmacia Biotech), treated with Pfu DNA polymerase (Stratagene) to fill-in restriction site overhangs, and transfected into competent E. coli cells. Transformed cells were selected on antibiotic-containing media, and individual colonies were picked and cultured overnight at 37°C in 384-well plates in LB/2x carb liquid media.

30 The cells were lysed, and DNA was amplified by PCR using Taq DNA polymerase (Amersham Pharmacia Biotech) and Pfu DNA polymerase (Stratagene) with the following parameters: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 72°C, 2 min; Step 5: steps 2, 3, and 4 repeated 29 times; Step 6: 72°C, 5 min; Step 7: storage at 4°C. DNA was quantified by PICOGREEN reagent (Molecular Probes) as described above. Samples with low DNA recoveries were reamplified

using the same conditions as described above. Samples were diluted with 20% dimethylsulfoxide (1:2, v/v), and sequenced using DYENAMIC energy transfer sequencing primers and the DYENAMIC DIRECT kit (Amersham Pharmacia Biotech) or the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems).

5 In like manner, full length polynucleotide sequences are verified using the above procedure or are used to obtain 5' regulatory sequences using the above procedure along with oligonucleotides designed for such extension, and an appropriate genomic library.

#### IX. Labeling and Use of Individual Hybridization Probes

Hybridization probes derived from SEQ ID NO:22-42 are employed to screen cDNAs, genomic  
10 DNAs, or mRNAs. Although the labeling of oligonucleotides, consisting of about 20 base pairs, is specifically described, essentially the same procedure is used with larger nucleotide fragments. Oligonucleotides are designed using state-of-the-art software such as OLIGO 4.06 software (National Biosciences) and labeled by combining 50 pmol of each oligomer, 250  $\mu$ Ci of [ $\gamma$ -<sup>32</sup>P] adenosine triphosphate (Amersham Pharmacia Biotech), and T4 polynucleotide kinase (DuPont NEN, Boston  
15 MA). The labeled oligonucleotides are substantially purified using a SEPHADEX G-25 superfine size exclusion dextran bead column (Amersham Pharmacia Biotech). An aliquot containing 10<sup>7</sup> counts per minute of the labeled probe is used in a typical membrane-based hybridization analysis of human genomic DNA digested with one of the following endonucleases: Ase I, Bgl II, Eco RI, Pst I, Xba I, or Pvu II (DuPont NEN).

20 The DNA from each digest is fractionated on a 0.7% agarose gel and transferred to nylon membranes (Nytran Plus, Schleicher & Schuell, Durham NH). Hybridization is carried out for 16 hours at 40°C. To remove nonspecific signals, blots are sequentially washed at room temperature under conditions of up to, for example, 0.1 x saline sodium citrate and 0.5% sodium dodecyl sulfate. Hybridization patterns are visualized using autoradiography or an alternative imaging means and  
25 compared.

#### X. Microarrays

The linkage or synthesis of array elements upon a microarray can be achieved utilizing photolithography, piezoelectric printing (ink-jet printing, See, e.g., Baldeschweiler, supra), mechanical microspotting technologies, and derivatives thereof. The substrate in each of the aforementioned  
30 technologies should be uniform and solid with a non-porous surface (Schena (1999), supra). Suggested substrates include silicon, silica, glass slides, glass chips, and silicon wafers. Alternatively, a procedure analogous to a dot or slot blot may also be used to arrange and link elements to the surface of a substrate using thermal, UV, chemical, or mechanical bonding procedures. A typical array may be produced using available methods and machines well known to those of ordinary skill in the art and may

contain any appropriate number of elements. (See, e.g., Schena, M. et al. (1995) *Science* 270:467-470; Shalon, D. et al. (1996) *Genome Res.* 6:639-645; Marshall, A. and J. Hodgson (1998) *Nat. Biotechnol.* 16:27-31.)

Full length cDNAs, Expressed Sequence Tags (ESTs), or fragments or oligomers thereof may 5 comprise the elements of the microarray. Fragments or oligomers suitable for hybridization can be selected using software well known in the art such as LASERGENE software (DNASTAR). The array elements are hybridized with polynucleotides in a biological sample. The polynucleotides in the biological sample are conjugated to a fluorescent label or other molecular tag for ease of detection. After hybridization, nonhybridized nucleotides from the biological sample are removed, and a 10 fluorescence scanner is used to detect hybridization at each array element. Alternatively, laser desorption and mass spectrometry may be used for detection of hybridization. The degree of complementarity and the relative abundance of each polynucleotide which hybridizes to an element on the microarray may be assessed. In one embodiment, microarray preparation and usage is described in detail below.

15 **Tissue or Cell Sample Preparation**

Total RNA is isolated from tissue samples using the guanidinium thiocyanate method and poly(A)<sup>+</sup>RNA is purified using the oligo-(dT)-cellulose method. Each poly(A)<sup>+</sup> RNA sample is reverse-transcribed using MMLV-reverse-transcriptase, 0.05 pg/ $\mu$ l oligo-(dT) primer (21mer), 1X first strand buffer, 0.03 units/ $\mu$ l RNase inhibitor, 500  $\mu$ M dATP, 500  $\mu$ M dGTP, 500  $\mu$ M dTTP, 40  $\mu$ M 20 dCTP, 40  $\mu$ M dCTP-Cy3 (BDS) or dCTP-Cy5 (Amersham Pharmacia Biotech). The reverse transcription reaction is performed in a 25 ml volume containing 200 ng poly(A)<sup>+</sup> RNA with GEMBRIGHT kits (Incyte). Specific control poly(A)<sup>+</sup> RNAs are synthesized by *in vitro* transcription from non-coding yeast genomic DNA. After incubation at 37°C for 2 hr, each reaction sample (one with Cy3 and another with Cy5 labeling) is treated with 2.5 ml of 0.5M sodium hydroxide and 25 incubated for 20 minutes at 85°C to stop the reaction and degrade the RNA. Samples are purified using two successive CHROMA SPIN 30 gel filtration spin columns (CLONTECH Laboratories, Inc. (CLONTECH), Palo Alto CA) and after combining, both reaction samples are ethanol precipitated using 1 ml of glycogen (1 mg/ml), 60 ml sodium acetate, and 300 ml of 100% ethanol. The sample is then dried to completion using a SpeedVAC (Savant Instruments Inc., Holbrook NY) and 30 resuspended in 14  $\mu$ l 5X SSC/0.2% SDS.

**Microarray Preparation**

Sequences of the present invention are used to generate array elements. Each array element is amplified from bacterial cells containing vectors with cloned cDNA inserts. PCR amplification uses primers complementary to the vector sequences flanking the cDNA insert. Array elements are

amplified in thirty cycles of PCR from an initial quantity of 1-2 ng to a final quantity greater than 5 µg. Amplified array elements are then purified using SEPHACRYL-400 (Amersham Pharmacia Biotech).

Purified array elements are immobilized on polymer-coated glass slides. Glass microscope  
5 slides (Corning) are cleaned by ultrasound in 0.1% SDS and acetone, with extensive distilled water washes between and after treatments. Glass slides are etched in 4% hydrofluoric acid (VWR Scientific Products Corporation (VWR), West Chester PA), washed extensively in distilled water, and coated with 0.05% aminopropyl silane (Sigma) in 95% ethanol. Coated slides are cured in a 110°C oven.

10 Array elements are applied to the coated glass substrate using a procedure described in US Patent No. 5,807,522, incorporated herein by reference. 1 µl of the array element DNA, at an average concentration of 100 ng/µl, is loaded into the open capillary printing element by a high-speed robotic apparatus. The apparatus then deposits about 5 nl of array element sample per slide.

Microarrays are UV-crosslinked using a STRATALINKER UV-crosslinker (Stratagene).  
15 Microarrays are washed at room temperature once in 0.2% SDS and three times in distilled water. Non-specific binding sites are blocked by incubation of microarrays in 0.2% casein in phosphate buffered saline (PBS) (Tropix, Inc., Bedford MA) for 30 minutes at 60°C followed by washes in 0.2% SDS and distilled water as before.

#### Hybridization

20 Hybridization reactions contain 9 µl of sample mixture consisting of 0.2 µg each of Cy3 and Cy5 labeled cDNA synthesis products in 5X SSC, 0.2% SDS hybridization buffer. The sample mixture is heated to 65°C for 5 minutes and is aliquoted onto the microarray surface and covered with an 1.8 cm<sup>2</sup> coverslip. The arrays are transferred to a waterproof chamber having a cavity just slightly larger than a microscope slide. The chamber is kept at 100% humidity internally by the addition of  
25 140 µl of 5X SSC in a corner of the chamber. The chamber containing the arrays is incubated for about 6.5 hours at 60°C. The arrays are washed for 10 min at 45°C in a first wash buffer (1X SSC, 0.1% SDS), three times for 10 minutes each at 45°C in a second wash buffer (0.1X SSC), and dried.

#### Detection

Reporter-labeled hybridization complexes are detected with a microscope equipped with an  
30 Innova 70 mixed gas 10 W laser (Coherent, Inc., Santa Clara CA) capable of generating spectral lines at 488 nm for excitation of Cy3 and at 632 nm for excitation of Cy5. The excitation laser light is focused on the array using a 20X microscope objective (Nikon, Inc., Melville NY). The slide containing the array is placed on a computer-controlled X-Y stage on the microscope and raster-scanned past the objective. The 1.8 cm x 1.8 cm array used in the present example is scanned with a  
35 resolution of 20 micrometers.

In two separate scans, a mixed gas multiline laser excites the two fluorophores sequentially. Emitted light is split, based on wavelength, into two photomultiplier tube detectors (PMT R1477, Hamamatsu Photonics Systems, Bridgewater NJ) corresponding to the two fluorophores. Appropriate filters positioned between the array and the photomultiplier tubes are used to filter the signals. The 5 emission maxima of the fluorophores used are 565 nm for Cy3 and 650 nm for Cy5. Each array is typically scanned twice, one scan per fluorophore using the appropriate filters at the laser source, although the apparatus is capable of recording the spectra from both fluorophores simultaneously.

The sensitivity of the scans is typically calibrated using the signal intensity generated by a cDNA control species added to the sample mixture at a known concentration. A specific location on 10 the array contains a complementary DNA sequence, allowing the intensity of the signal at that location to be correlated with a weight ratio of hybridizing species of 1:100,000. When two samples from different sources (e.g., representing test and control cells), each labeled with a different fluorophore, are hybridized to a single array for the purpose of identifying genes that are differentially expressed, the calibration is done by labeling samples of the calibrating cDNA with the two 15 fluorophores and adding identical amounts of each to the hybridization mixture.

The output of the photomultiplier tube is digitized using a 12-bit RTI-835H analog-to-digital (A/D) conversion board (Analog Devices, Inc., Norwood MA) installed in an IBM-compatible PC computer. The digitized data are displayed as an image where the signal intensity is mapped using a linear 20-color transformation to a pseudocolor scale ranging from blue (low signal) to red (high signal). The data is also analyzed quantitatively. Where two different fluorophores are excited and measured simultaneously, the data are first corrected for optical crosstalk (due to overlapping emission spectra) between the fluorophores using each fluorophore's emission spectrum.

A grid is superimposed over the fluorescence signal image such that the signal from each spot is centered in each element of the grid. The fluorescence signal within each element is then integrated 25 to obtain a numerical value corresponding to the average intensity of the signal. The software used for signal analysis is the GEMTOOLS gene expression analysis program (Incyte).

## XI. Complementary Polynucleotides

Sequences complementary to the PRTS-encoding sequences, or any parts thereof, are used to detect, decrease, or inhibit expression of naturally occurring PRTS. Although use of oligonucleotides 30 comprising from about 15 to 30 base pairs is described, essentially the same procedure is used with smaller or with larger sequence fragments. Appropriate oligonucleotides are designed using OLIGO 4.06 software (National Biosciences) and the coding sequence of PRTS. To inhibit transcription, a complementary oligonucleotide is designed from the most unique 5' sequence and used to prevent promoter binding to the coding sequence. To inhibit translation, a complementary oligonucleotide is 35 designed to prevent ribosomal binding to the PRTS-encoding transcript.

**XII. Expression of PRTS**

Expression and purification of PRTS is achieved using bacterial or virus-based expression systems. For expression of PRTS in bacteria, cDNA is subcloned into an appropriate vector containing an antibiotic resistance gene and an inducible promoter that directs high levels of cDNA transcription.

5 Examples of such promoters include, but are not limited to, the *trp-lac* (*tac*) hybrid promoter and the T5 or T7 bacteriophage promoter in conjunction with the *lac* operator regulatory element. Recombinant vectors are transformed into suitable bacterial hosts, e.g., BL21(DE3). Antibiotic resistant bacteria express PRTS upon induction with isopropyl beta-D-thiogalactopyranoside (IPTG). Expression of PRTS in eukaryotic cells is achieved by infecting insect or mammalian cell lines with

10 recombinant *Autographica californica* nuclear polyhedrosis virus (AcMNPV), commonly known as baculovirus. The nonessential polyhedrin gene of baculovirus is replaced with cDNA encoding PRTS by either homologous recombination or bacterial-mediated transposition involving transfer plasmid intermediates. Viral infectivity is maintained and the strong polyhedrin promoter drives high levels of cDNA transcription. Recombinant baculovirus is used to infect *Spodoptera frugiperda* (Sf9) insect

15 cells in most cases, or human hepatocytes, in some cases. Infection of the latter requires additional genetic modifications to baculovirus. (See Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945.)

In most expression systems, PRTS is synthesized as a fusion protein with, e.g., glutathione S-transferase (GST) or a peptide epitope tag, such as FLAG or 6-His, permitting rapid, single-step,

20 affinity-based purification of recombinant fusion protein from crude cell lysates. GST, a 26-kilodalton enzyme from *Schistosoma japonicum*, enables the purification of fusion proteins on immobilized glutathione under conditions that maintain protein activity and antigenicity (Amersham Pharmacia Biotech). Following purification, the GST moiety can be proteolytically cleaved from PRTS at specifically engineered sites. FLAG, an 8-amino acid peptide, enables immunoaffinity purification

25 using commercially available monoclonal and polyclonal anti-FLAG antibodies (Eastman Kodak). 6-His, a stretch of six consecutive histidine residues, enables purification on metal-chelate resins (QIAGEN). Methods for protein expression and purification are discussed in Ausubel (1995, *supra*, ch. 10 and 16). Purified PRTS obtained by these methods can be used directly in the assays shown in Examples XVI, XVII, XVIII, and XIX, where applicable.

**30 XIII. Functional Assays**

PRTS function is assessed by expressing the sequences encoding PRTS at physiologically elevated levels in mammalian cell culture systems. cDNA is subcloned into a mammalian expression vector containing a strong promoter that drives high levels of cDNA expression. Vectors of choice include PCMV SPORT (Life Technologies) and PCR3.1 (Invitrogen, Carlsbad CA), both of which

contain the cytomegalovirus promoter. 5-10  $\mu$ g of recombinant vector are transiently transfected into a human cell line, for example, an endothelial or hematopoietic cell line, using either liposome formulations or electroporation. 1-2  $\mu$ g of an additional plasmid containing sequences encoding a marker protein are co-transfected. Expression of a marker protein provides a means to distinguish 5 transfected cells from nontransfected cells and is a reliable predictor of cDNA expression from the recombinant vector. Marker proteins of choice include, e.g., Green Fluorescent Protein (GFP; Clontech), CD64, or a CD64-GFP fusion protein. Flow cytometry (FCM), an automated, laser optics-based technique, is used to identify transfected cells expressing GFP or CD64-GFP and to evaluate the apoptotic state of the cells and other cellular properties. FCM detects and quantifies the uptake of 10 fluorescent molecules that diagnose events preceding or coincident with cell death. These events include changes in nuclear DNA content as measured by staining of DNA with propidium iodide; changes in cell size and granularity as measured by forward light scatter and 90 degree side light scatter; down-regulation of DNA synthesis as measured by decrease in bromodeoxyuridine uptake; alterations in expression of cell surface and intracellular proteins as measured by reactivity with specific antibodies; 15 and alterations in plasma membrane composition as measured by the binding of fluorescein-conjugated Annexin V-protein to the cell surface. Methods in flow cytometry are discussed in Ormerod, M.G. (1994) Flow Cytometry, Oxford, New York NY.

The influence of PRTS on gene expression can be assessed using highly purified populations of 20 cells transfected with sequences encoding PRTS and either CD64 or CD64-GFP. CD64 and CD64-GFP are expressed on the surface of transfected cells and bind to conserved regions of human immunoglobulin G (IgG). Transfected cells are efficiently separated from nontransfected cells using magnetic beads coated with either human IgG or antibody against CD64 (DYNAL, Lake Success NY). mRNA can be purified from the cells using methods well known by those of skill in the art. Expression 25 of mRNA encoding PRTS and other genes of interest can be analyzed by northern analysis or microarray techniques.

#### XIV. Production of PRTS Specific Antibodies

PRTS substantially purified using polyacrylamide gel electrophoresis (PAGE; see, e.g., Harrington, M.G. (1990) *Methods Enzymol.* 182:488-495), or other purification techniques, is used to 30 immunize rabbits and to produce antibodies using standard protocols.

Alternatively, the PRTS amino acid sequence is analyzed using LASERGENE software (DNASTAR) to determine regions of high immunogenicity, and a corresponding oligopeptide is synthesized and used to raise antibodies by means known to those of skill in the art. Methods for selection of appropriate epitopes, such as those near the C-terminus or in hydrophilic regions are well described in the art. (See, e.g., Ausubel, 1995, supra, ch. 11.)

Typically, oligopeptides of about 15 residues in length are synthesized using an ABI 431A peptide synthesizer (Applied Biosystems) using FMOC chemistry and coupled to KLH (Sigma-Aldrich, St. Louis MO) by reaction with N-maleimidobenzoyl-N-hydroxysuccinimide ester (MBS) to increase immunogenicity. (See, e.g., Ausubel, 1995, *supra*.) Rabbits are immunized with the 5 oligopeptide-KLH complex in complete Freund's adjuvant. Resulting antisera are tested for antipeptide and anti-PRTS activity by, for example, binding the peptide or PRTS to a substrate, blocking with 1% BSA, reacting with rabbit antisera, washing, and reacting with radio-iodinated goat anti-rabbit IgG.

#### XV. Purification of Naturally Occurring PRTS Using Specific Antibodies

Naturally occurring or recombinant PRTS is substantially purified by immunoaffinity 10 chromatography using antibodies specific for PRTS. An immunoaffinity column is constructed by covalently coupling anti-PRTS antibody to an activated chromatographic resin, such as CNBr-activated SEPHAROSE (Amersham Pharmacia Biotech). After the coupling, the resin is blocked and washed according to the manufacturer's instructions.

Media containing PRTS are passed over the immunoaffinity column, and the column is washed 15 under conditions that allow the preferential absorbance of PRTS (e.g., high ionic strength buffers in the presence of detergent). The column is eluted under conditions that disrupt antibody/PRTS binding (e.g., a buffer of pH 2 to pH 3, or a high concentration of a chaotropic, such as urea or thiocyanate ion), and PRTS is collected.

#### XVI. Identification of Molecules Which Interact with PRTS

PRTS, or biologically active fragments thereof, are labeled with  $^{125}\text{I}$  Bolton-Hunter reagent. 20 (See, e.g., Bolton A.E. and W.M. Hunter (1973) Biochem. J. 133:529-539.) Candidate molecules previously arrayed in the wells of a multi-well plate are incubated with the labeled PRTS, washed, and any wells with labeled PRTS complex are assayed. Data obtained using different concentrations of PRTS are used to calculate values for the number, affinity, and association of PRTS with the candidate 25 molecules.

Alternatively, molecules interacting with PRTS are analyzed using the yeast two-hybrid system as described in Fields, S. and O. Song (1989) Nature 340:245-246, or using commercially available kits based on the two-hybrid system, such as the MATCHMAKER system (Clontech).

PRTS may also be used in the PATHCALLING process (CuraGen Corp., New Haven CT) 30 which employs the yeast two-hybrid system in a high-throughput manner to determine all interactions between the proteins encoded by two large libraries of genes (Nandabalan, K. et al. (2000) U.S. Patent No. 6,057,101).

#### XVII. Demonstration of PRTS Activity

Protease activity is measured by the hydrolysis of appropriate synthetic peptide substrates conjugated with various chromogenic molecules in which the degree of hydrolysis is quantified by spectrophotometric (or fluorometric) absorption of the released chromophore (Beynon, R.J. and J.S. Bond (1994) Proteolytic Enzymes: A Practical Approach, Oxford University Press, New York NY, pp.25-55). Peptide substrates are designed according to the category of protease activity as endopeptidase (serine, cysteine, aspartic proteases, or metalloproteases), aminopeptidase (leucine aminopeptidase), or carboxypeptidase (carboxypeptidases A and B, procollagen C-proteinase). Commonly used chromogens are 2-naphthylamine, 4-nitroaniline, and furylacrylic acid. For example, arginine- $\beta$ -naphthylamide can be used as a substrate for SEQ ID NO:3 (Fukasawa, K.M. et al. (1996) J. Biol. Chem. 271:30731-30735) and 4-phenylazobenzoyloxycarbonyl-Pro-Leu-Gly-Pro-D-Arg can be used as a substrate for SEQ ID NO:4. In an alternative example, a substrate for SEQ ID NO:9 would be 7-amino-4-trifluoromethyl coumarin-Phe-Pro-AFC. Assays are performed at ambient temperature and contain an aliquot of the enzyme and the appropriate substrate in a suitable buffer. Reactions are carried out in an optical cuvette, and the increase/decrease in absorbance of the chromogen released during hydrolysis of the peptide substrate is measured. The change in absorbance is proportional to the enzyme activity in the assay.

An alternate assay for ubiquitin hydrolase activity measures the hydrolysis of a ubiquitin precursor. The assay is performed at ambient temperature and contains an aliquot of PRTS and the appropriate substrate in a suitable buffer. For SEQ ID NO:1, chemically synthesized human ubiquitin-valine may be used as substrate. Cleavage of the C-terminal valine residue from the substrate is monitored by capillary electrophoresis (Franklin, K. et al. (1997) Anal. Biochem. 247:305-309). Alternatively, the ubiquitin protease activity of SEQ ID NO:5 can be measured using the method of Sloper-Mould et al. ((1999) J. Biol. Chem. 274:26878-26884). Aliquots of PRTS are incubated with 5  $\mu$ l [ $^{35}$ S]-labeled ubiquitin-GST fusion substrate for 1 hour at 37 °C in an appropriate buffer. Samples are resolved by electrophoresis on a 12% SDS-PAGE gel. Ubiquitin cleavage is monitored by fluorography of the gel.

Alternatively, the activity of SEQ ID NO:10, for example, can be measured by the method of Colige et al. (1999, J. Biol. Chem. 270:16724-16730). An aliquot of PRTS is incubated with amino procollagen type I substrate radioactively labeled only in the propeptide in a 250  $\mu$ l reaction containing 50 mM sodium cacodylate, pH 7.5, 200 mM KCl, 2 mM CaCl, 2.5 mM NEM, 0.5 mM PMSF, and 0.02% Brij (standard assay buffer). After 16 h at 26 °C, the reaction is stopped by adding 50  $\mu$ l of EDTA solution (0.2 M EDTA, pH 8, 0.5% SDS, 0.5 M DTT) and 300  $\mu$ l of 99% ethanol. The samples are kept for 30 min at 4 °C and centrifuged for 30 min at 9500 g. Collagen and uncleaved

radioactive pN-collagen substrate are pelleted, whereas the freed amino propeptides remained in solution. An aliquot of the supernatant is assayed by liquid scintillation spectrometry.

In the alternative, an assay for protease activity takes advantage of fluorescence resonance energy transfer (FRET) that occurs when one donor and one acceptor fluorophore with an appropriate spectral overlap are in close proximity. A flexible peptide linker containing a cleavage site specific for PRTS is fused between a red-shifted variant (RSGFP4) and a blue variant (BFP5) of Green Fluorescent Protein. This fusion protein has spectral properties that suggest energy transfer is occurring from BFP5 to RSGFP4. When the fusion protein is incubated with PRTS, the substrate is cleaved, and the two fluorescent proteins dissociate. This is accompanied by a marked decrease in energy transfer which is quantified by comparing the emission spectra before and after the addition of PRTS (Mitra, R.D. et al. (1996) Gene 173:13-17). This assay can also be performed in living cells. In this case the fluorescent substrate protein is expressed constitutively in cells and PRTS is introduced on an inducible vector so that FRET can be monitored in the presence and absence of PRTS (Sagot, I. et al. (1999) FEBS Lett. 447:53-57).

15 In yet another alternative, an assay for PRTS dipeptidase activity measures the hydrolysis activity of PRTS on a variety of dipeptides such as leukotriene D4 (Kozak, E. and S. Tate (1982) J. Biol. Chem. 257:6322-6327), or hydrolysis of the beta-lactam ring of antibiotics such as penum and carbapenem (Campbell et al., (1984) J. Biol. Chem. 259:14586-14590).

### XVIII. Identification of PRTS Substrates

20 Phage display libraries can be used to identify optimal substrate sequences for PRTS. A random hexamer followed by a linker and a known antibody epitope is cloned as an N-terminal extension of gene III in a filamentous phage library. Gene III codes for a coat protein, and the epitope will be displayed on the surface of each phage particle. The library is incubated with PRTS under proteolytic conditions so that the epitope will be removed if the hexamer codes for a PRTS cleavage site. An antibody that recognizes the epitope is added along with immobilized protein A. Uncleaved phage, which still bear the epitope, are removed by centrifugation. Phage in the supernatant are then amplified and undergo several more rounds of screening. Individual phage clones are then isolated and sequenced. Reaction kinetics for these peptide substrates can be studied using an assay in Example XVII, and an optimal cleavage sequence can be derived (Ke, S.H. et al. (1997) J. Biol. Chem. 272:16603-16609).

25

30

To screen for in vivo PRTS substrates, this method can be expanded to screen a cDNA expression library displayed on the surface of phage particles (T7SELECT 10-3 Phage display vector, Novagen, Madison WI) or yeast cells (pYD1 yeast display vector kit, Invitrogen, Carlsbad CA). In this case, entire cDNAs are fused between Gene III and the appropriate epitope.

**XIX. Identification of PRTS Inhibitors**

Compounds to be tested are arrayed in the wells of a multi-well plate in varying concentrations along with an appropriate buffer and substrate, as described in the assays in Example XVII. PRTS activity is measured for each well and the ability of each compound to inhibit PRTS activity can be 5 determined, as well as the dose-response kinetics. This assay could also be used to identify molecules which enhance PRTS activity.

In the alternative, phage display libraries can be used to screen for peptide PRTS inhibitors. Candidates are found among peptides which bind tightly to a protease. In this case, multi-well plate wells are coated with PRTS and incubated with a random peptide phage display library or a cyclic 10 peptide library (Koivunen, E. et al. (1999) Nat. Biotechnol. 17:768-774). Unbound phage are washed away and selected phage amplified and rescreened for several more rounds. Candidates are tested for PRTS inhibitory activity using an assay described in Example XVII.

Various modifications and variations of the described methods and systems of the invention will 15 be apparent to those skilled in the art without departing from the scope and spirit of the invention.

Although the invention has been described in connection with certain embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention which are obvious to those skilled in molecular biology or related fields are intended to be within the scope of the following 20 claims.

Table 1

Incyte Project ID	Polypeptide SEQ ID NO:	Incyte Polypeptide ID	Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID
275791	1	275791CD1	22	275791CB1
1389845	2	1389845CD1	23	1389845CB1
1726609	3	1726609CD1	24	1726609CB1
4503848	4	4503848CD1	25	4503848CB1
5544089	5	5544089CD1	26	5544089CB1
7474081	6	7474081CD1	27	7474081CB1
5281209	7	5281209CD1	28	5281209CB1
2256251	8	2256251CD1	29	2256251CB1
7160544	9	7160544CD1	30	7160544CB1
7477386	10	7477386CD1	31	7477386CB1
7473089	11	7473089CD1	32	7473089CB1
7604035	12	7604035CD1	33	7604035CB1
3473847	13	3473847CD1	34	3473847CB1
3750004	14	3750004CD1	35	3750004CB1
4904126	15	4904126CD1	36	4904126CB1
71268415	16	71268415CD1	37	71268415CB1
7473301	17	7473301CD1	38	7473301CB1
7473308	18	7473308CD1	39	7473308CB1
7478021	19	7478021CD1	40	7478021CB1
4333459	20	4333459CD1	41	4333459CB1
6817347	21	6817347CD1	42	6817347CB1

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO:	Probability score	GenBank Homolog
1	275791CD1	g9971757	1.00E-49	Ubiquitin-specific processing protease [Homo sapiens]
2	1389845CD1	g1143194	1.30E-46	Biosatin [Homo sapiens]
3	1726609CD1	g10719660	0	RNPEPLike protein [Homo sapiens] (Horikawa, Y. et al. (2000) Nat. Genet. 26:163-175.)
		g1754515	3.30E-96	aminopeptidase-B [Rattus norvegicus] (Prieto, I. et al. (1998) Horm. Metab. Res. 30:246-248.)
4	4503848CD1	g1783122	0	endopeptidase 24.16 type M1 [Sus scrofa]
5	5544089CD1	g5410230	5.20E-43	ubiquitin-specific protease 3 [Homo sapiens]
6	7474081CD1	g603903	2.90E-33	Trypsinogen [Gallus gallus]
7	5281209CD1	g11071729	0	putative dipeptidase [Homo sapiens] dipeptidase precursor [Homo sapiens] (Satoh, S. et al. (1993) Biochim. Biophys. Acta 1172:181-183.)
8	2256251CD1	g6103629	3.90E-166	transmembrane tryptase [Homo sapiens] (Wong, G.W. et al. (1999) J. Biol. Chem. 274:30784-30793.)
9	7160544CD1	g11095188	0	dipeptidyl peptidase 8 [Homo sapiens] (Abbott, C.A. et al. (2000) Eur. J. Biochem. 267:6140-6150.)
		g1753197	6.80E-64	dipeptidyl peptidase IV [Stenotrophomonas maltophilia] (Mentlein, R. (1999) Regul. Pept. 85:9-24; Kahne, T. Int. J. Mol. Med. (1999) 4:3-15.)
10	7477386CD1	g1865716	0	Procollagen I N-Proteinase [Bos taurus] (Colige, A. et al. (1999) Am. J. Hum. Genet. 65:308-317.)
11	7473089CD1	g7768706	3.60E-255	metalloprotease with thrombospondin type 1 motifs [Homo sapiens] (Vazquez, F. et al. (1999) J. Biol. Chem. 274:23349-23357.)
12	7604035CD1	g6164595	4.70E-68	Lacumin [Manduca sexta]

Table 2 (cont.)

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO:	Probability score	GenBank Homolog
13	3473847CD1	g217172	9.20E-50	aqualysin precursor (aa 1 to 513) [Thermus aquaticus]
14	3750004CD1	g5923786	4.30E-51	zinc metalloprotease ADAMTS6 [Homo sapiens]
15	4904126CD1	g186286	3.90E-40	interleukin 1-beta convertase [Homo sapiens]
16	71268415CD1	g6651071	0	(Cerretti, D.P. et al. (1992) Science 256:97-100.) disintegrin and metalloproteinase domain 19 [Homo sapiens]
17	7473301CD1	g6002714	5.10E-94	(Inoue, D. et al. (1998) J. Biol. Chem. 273:4180-4187.) membrane-type serine protease 1 [Homo sapiens]
18	7473308CD1	g1552517	6.60E-77	(Takeuchi, T. et al. (1999) Proc. Natl. Acad. Sci. USA 96:11054-11061.)
19	7478021CD1	g3211705	5.60E-189	trypsinogen E [Homo sapiens]
20	4333459CD1	g1754714	2.30E-67	(Yang, M. (1997) J. Biol. Chem. 272:13527-13533.) Orviductin [Xenopus laevis]
21	6817347CD1	g7673618	5.10E-283	(Lindsay, L.L. et al. (1999) Biol. Reprod. 60:989-995.) ubiquitin specific protease [Mus musculus]

Table 3

SEQ NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
1	275791CD1	232	T15 T17 S23 S43 S71 T90 S93 S100 S107 S111 T122 T174 S190 T10 S141 S190 T213 T227	N98 N99	Ubiquitin-carboxyl-terminal hydrolase Family 2 signature 2 Uch_2_2: Ubiquitin carboxyl-terminal hydrolase Family 2 signature 2 UCH-2: L138-H203 Ubiquitin carboxyl-terminal hydrolase Family 2 signature 2 BL00972: Y142- D166, K169-S190	MOTIFS
2	1389845CD1	365	S120 S187 S225 S253 S82 T31 T37 T42 Y283		TRYPSIN DM00018 A57014  45-284: I123-Q314 BLAST_DOMO PROTEASE SERINE PRECURSOR SIGNAL HYDROLASE ZYMOGEN GLYCOPROTEIN FAMILY MULTIGENE FACTOR PD000046: I123-Q314 Serine proteases, trypsin family BL00134A: C148-C164	BLAST_PRODOM BLIMPS_BLOCKS BLIMPS_PRINTS

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Potential Glycosylation Domains and Motifs	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
3	1726609CD1	416	S244 S283 S30 S370 S408 T389 T59 T78 T87	N203 N413 N57	do HYDROLASE; LEUKOTRIENE; A-4 ; ZINC ; DM08707   P19602   7-609 : M1-P354 AMINOPEPTIDASE HYDROLASE METALLOPROTEASE BLAST_PRODOM ZINC N GLYCOPROTEIN PROTEIN TRANSMEMBRANE SIGNAL ANCHOR MEMBRANE PD001134: R4-S177		BLAST_DOMO
4	4503848CD1	714	S124 S140 S147 S179 S200 S206 S226 S333 S551 S556 S592 T114 T133 T244 T252 T270 T308 T318 T322 T376 T406 T432 T528 T585 T602 T69 Y175 Y249 Y505	N425 N485 N601	do ZINC; METALLOPEPTIDASE; NEUTRAL; OLIGO-PEPTIDASE DM01184   Q02038   36-702 : A46-A713	SPScan.	BLIMPS_BLOCKSBLS

Table 3 (cont.)

SEQ NO:	Incyte ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
5	5544089CD1	367	S108 S161 S197 S203 S235 S266 S361 S49 T180 T263 T316 T331	N139 N142 N308	UBIQUITIN CARBOXYL-TERMINAL HYDROLASES FAMILY 2 DM00659   P40818   782-1103: L36-L328	BLAST_DOMO
6	7474081CD1	235	S134 S143 S159 S171 S226 T231	N90	Ubiquitin carboxyl-terminal hydrolase Family BL00972: Y74-L83, V120-C134, Y274-N298, G301-K322	BLIMPS_BLOCKS
					Ubiquitin carboxyl-terminal hydrolase Family UCH-2: E270-Q332	HMMER_PFAM
					Ubiquitin carboxyl-terminal hydrolase Family Uch_2_2: Y274-Y292	MOTIFS
					SPScan	
					TRYPSIN DM00018   S55065   26-244: A27-T231	BLAST_DOMO
					Serine proteases, trypsin family signature BL00134: C40-C56, V215-I228	BLIMPS_BLOCKS
					Type I fibronectin domain BL01253: C40-P53, I197-R231	BLIMPS_BLOCKS
					Kringle domain proteins BL00021: S159-Q164, C40-Y57	BLIMPS_BLOCKS
					CHYMOTRYPSIN SERINE PROTEASE PR00722A: V41-C56, A95-A109	BLIMPS_PRINTS
					Serine proteases, trypsin family, active PROFILESCAN site trypsin_his.prf: S35-T76	
					Trypsin: G42-V178, G216-I228	HMMER_PFAM
					Leucine_Zipper: L44-L65	MOTIFS
					signal_cleavage: M1-S19	SPScan

Table 3 (cont.)

SEQ NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
7	5281209CD1	488	S13 T74 S186 S233 T363 T456 T4 T34 T125 S170 S172 T178 S249 T337 S387 S389 S419 T447	N119 N184 N243 N334	Renal dipeptidase proteins BL00869 : P92-L247, E280-R412, S415-N457 DIPEPTIDASE MICROSMAL PRECURSOR MDP HYDROLASE MICROSOME SIGNAL GPI-ANCHOR GLYCOPROTEIN ZINC PD005626 : S143-E450 RENAL DIPEPTIDASE DM02775 : T77-R410 Renal dipeptidase : V195-R217	BLIMPS-BLOCKS BLAST-PRODOM BLAST-DOMO MOTIFS
8	2256251CD1	346	S203 S210 S266 S45 S79 T131 T147 T216	N110	TRYPSIN DM00018   P15944   31-270 : I63-I294 PROTEASE SERINE PRECURSOR SIGNAL HYDROLASE ZYMOGEN GLYCOPROTEIN FAMILY MULTIGENE FACTOR PD00046: L156-I290, I63-F178, N288-F314, P274-P305	HMMER-PFAM HMMER SPSCAN BLAST_DOMO BLAST_PRODOM

Table 3 (cont.)

SEQ NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Potential Glycosylation Domains and Motifs	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
9	7160544CD1	882	S115 S133 S293 S312 S412 S443 S479 S530 S587 S588 S723 S804 S850 T227 T234 T307 T326 T499 T52 T551 T594; T603 T615 T776 Y315 Y36 Y55 Y555 Y844			PROLYL ENDOPEPTIDASE FAMILY SERINE DM02461   P27487   192-765 : F488-E870, G251-E370	BLAST_DOMO

					Dipeptidyl IV HYDROLASE PROTEASE SERINE PEPTIDASE DIPEPTIDASE TRANSMEMBRANE GLYCOPROTEIN PROTEIN PDD03048 : L744-E870	BLAST_PRODOM
					DIPEPTIDYL IV HYDROLASE PROTEASE SERINE PEPTIDASE DIPEPTIDASE TRANSMEMBRANE GLYCOPROTEIN PROTEIN PDD03086 : Y423- V661, I212-T326	BLAST_PRODOM
					Prolyl endopeptidase Family BL00708 : G501-I513, Q714-L744	BLIMPS_BLOCKS
					Dipeptidyl peptidase IV PF00930: I498- R508, F756-P783, R808-L828	BLIMPS_PFAM
					PROLYL OLIGOPEPTIDASE SERINE PROTEASE PR00862: P647-F665, G737-R757	BLIMPS_PRINTS
					Prolyl oligopeptidase family Peptidase_S9 : R672-L744	HMMER_PFAM
					Dipeptidyl peptidase IV DPPIV_N_term: M88-N663	HMMER_PFAM

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
10	7477386CD1	1189	S132 S169 S200 S32 S323 S350 S445 S480 S511 S626 S675 S699 S798 S1064 T247 T362 T521 T612 T718 T777 T946 T986 T1104 Y552	N109 N478 N944	do ZINC; METALLOPEPTIDASE; NEUTRAL ATROLYSIN DM00368  Q05910 189-395:I261-P463 THROMBOSPONDIN TYPE 1 REPEAT DM00275  P07996 477-540: D555-C604 PROTEIN PROCOLLAGEN THROMBOSPONDIN MOTIFS NPROTEINASE C02B4.1 A DISINTTEGRIN METALLOPROTEASE PD013511: L474-E549 PROTEIN PROCOLLAGEN THROMBOSPONDIN MOTIFS NPROTEINASE A DISINTTEGRIN METALLOPROTEASE WITH ADAMTS1 PD011654: Q647-C716 PROCOLLAGEN C37C3.6 SERINE PROTEASE INHIBITOR PD007018: W854-Q974, W914-C1029, W558-K623 METALLOPROTEASE PRECURSOR HYDROLASE SIGNAL ZINC VENOM CELL PROTEIN TRANSMEMBRANE ADHESION PD000791:P256-P463 Neutral zinc metallopeptidase B1.00142: T398-N408 signal_peptide: M1-A22 Reprolysin family propeptide Pep_M12B_propep: R120-V240 Reprolysin (M12B) Family zinc metalloprotease Reprolysin: I261-P463 Thrombospondin type 1 domain tsp_1: A973-C1024, S559-C609, Y852-C909, W914-C971 signal_cleavage: M1-G23	BLAST_DOMO BLAST_DOMO BLAST_DOMO BLAST_DOMO BLAST_DOMO BLAST_DOMO BLAST_DOMO BLAST_DOMO BLIMPS_BLOCKS HMMER HMMER_PFAM HMMER_PFAM HMMER_PFAM SPSCAN

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
11	7473089CD1	952	S19 S203 S207 S303 S346 S432 S492 S575 S578 S611 S666 S682 S708 S745 S919 T171 T288 T317 T325 T337 T359 T471 T594 T687 T765	N141 N591 N623 N680	d6_ZINC; METALLOPEPTIDASE; NEUTRAL; ATROLYSTIN DM00368   JJC2550   1-201:R218-P427 PROTEIN PROCOLLAGEN THROMBOSPONDIN MOTIFS NPROTEINASE A DISINTEGRIN METALLOPROTEASE WITH ADAMTS1 PD014161: K684-E804	BLAST_DOMO BLAST_PRODOM

Table 3 (cont.)

SEQ NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
12	7604035CD1	898	S187 S188 S258 S268 S285 S415 S467 S547 S696 S796 S819 S851 S892 T106 T198 T35 T434 T483 T492 T5	N3 N490 N773	PROCOLLAGEN C37C3 . 6 SERINE PROTEASE INHIBITOR ALTERNATIVE PD007018: Y726-C841, W786-A874, Y667-C778, W50-Q72, S368-Q383	BLAST_PRODOM
					PROTEIN PROCOLLAGEN THROMBOSPONDIN MOTIFS NPROTEINASE A DISINTTEGRIN METALLOPROTEASE WITH ADAMTS1 PD011654: Q416-C484	BLAST_PRODOM
13	3473847CD1	631	S117 S160 S174 S185 S188 S268 S28 S30 S358 S431 S503 S605 T142 T33 T346 T512 T606	N472	PROTEIN PROCOLLAGEN THROMBOSPONDIN MOTIFS NPROTEINASE A DISINTTEGRIN METALLOPROTEASE WITH ADAMTS1 PD014161: R485-I599  signal peptide: M1-D24 HMMER Thrombospondin type 1 domain tsp_1: G48-R87, W727-C783, E787-C841  signal cleavage: M1-D24 SPSCAN	BLAST_PRODOM  HMMER_HMMER_PFAM  SPSCAN

Table 3 (cont.)

SEQ NO:	Incyte ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
14	3750004CD1	470	S454 T104 T276 T386 T464	N182 N203 C81	Thrombospondin type 1 domain tsp_1: T34-C119-C185 PROTEIN PROCOLLAGEN THROMBOSPONDIN MOTIFS NPROTEINASE A DISINTEGRIN METALLOPROTEASE WITH ADAMTS1 PDD011654:	HMMER_PFAM BLAST_PRODOM
15	4904126CD1	110	S16 S36 T100 T49	N47	signal peptide: M1-G29 signal cleavage: M1-G24 Caspase recruitment domain CARD: A2-A91 INTERLEUKIN-1 BETA CONVERTING ENZYME FAMILY HISTIDINE DM07463   P29466   1-122 : M1-S110	HMMER_SPScan HMMER_PFAM BLAST_DOMO

Table 3 (cont.)

SEQ NO:	Incyte ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
16	71268415CD1	879	S132 S14 S208 S288 S571 S711 S747 S754 S755 S827 T106 T118 T29 T30 T373 T412 T42 T428 T444 T55 T688 Y167 Y39	N368 N371 N569 N68	Reprolysin Family Propertide Pep_M12B propep: R8-K119 Reprolysin (M12B) family zinc metalloprotease Reprolysin: K134-P332 Disintegrin: E349-Q424 do-ZINC; METALLOPEPTIDASE; NEUTRAL; ATROLYSIN; DM00368 S60257 204-414; K126-D333 do ZINC; REGULATED; EPIDYMAL; NEUTRAL; DM00591 S60257 492-628; F410-L549 MELTRIN, BETA METALLOPROTEASE	HMMER_PFMAM HMMER_PFMAM HMMER_PFMAM HMMER_PFMAM BLAST_DOMO	

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					Disintegrins signature E360-P419	PROFILESCAN
					Neutral zinc metallopeptidases, zinc-binding region signature Zinc-protease.prf: S249-G301	PROFILESCAN
					Transmembrane domain: V624-Y645	HMMER
					Zn binding region Zinc_Protease: T266-E275	MOTIFS
17	7473301CD1	850	S100 S275 S295 S358 S429 S448 S470 S474 S495 S536 S596 S64	N19 N210 N422 N486 N533 N559 N568	TRYPSIN FAMILY SERINE PROTEASE trypsin: I613-I842	HMMER_PFAM
			S787 S802 S807		Low-density lipoprotein receptor domain Ldl_recept_a: Q489-S527, P530-Q562, I564-C603	HMMER_PFAM
			T117 T250 T312		TRYPSIN DM00018   P98072   800-1033 : R612-V846	BLAST_DOMO
			T348 T382 T404		PROTEASE SERINE PRECURSOR SIGNAL HYDROLASE ZYMOGEN GLYCOPROTEIN FAMILY MULTIGENE FACTOR PD00046: Q675-I842, I613-G809	BLAST_PRODOM
			T426 T570 T714		Serine proteases, trypsin family BL00134: C638-C654, D791-T814, P829-I842	BLIMPS_BLOCKS
			T777		Type I fibronectin domain BL01253: C638-BLIMPS_BLOCKS A651, R790-C803, W811-Y845	BLIMPS_BLOCKS
					Kringle domain proteins BL00021: I722-G743, L801-I842, C638-F655	BLIMPS_BLOCKS
					LOW DENSITY LIPOPROTEIN RECEPTOR DOMAIN PR00261: G501-E522	BLIMPS_PRINTS
					CHYMOTRYPSIN SERINE PROTEASE PR00722: G639-C654, T697-W711, R790-S802	BLIMPS_PRINTS

Table 3 (cont.)

SEQ NO:	Incyte ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					Trypsin family serine protease active sites trypsin_his.prf: L630-K679 trypsin_ser.prf: I776-R825 transmembrane domain: I77-L95	PROFILESCAN
18	7473308CD1	254	S136 S14 S153 S195 S227 T230 T249		Trypsin family serine protease active sites trypsin_His_L649-C654 Trypsin_Ser_D791-S802	HMMER MOTIFS
					TRYPSIN FAMILY SERINE PROTEASE trypsin: I21-Q183 CHYMATRYPSIN SERINE PROTEASE FAMILY PR00722B: T89-A103	HMMER_PFAM BLIMPS_PRINTS
					TRYPSIN DM00018   P07478   24-242 : I21-Q183 PROTEASE SERINE PRECURSOR SIGNAL HYDROLASE ZYMOGEN GLYCOPROTEIN FAMILY	BLAST_DOMO BLAST_PRODOM
					MULTIGENE FACTOR PD000046: G23-Q183	

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
19	7478021CD1	568	S142 S145 S153 S172 S177 S190 S244 S316 S34 S420 S448 S552 T209 T22 T293 T334 T401 T427 T489 T79 Y509	N371	Matixin Peptidase_M10: F56-T266 Hemopexin domain: F332-T390, I393-S448, L450-Q498, I505-K548 MATRIXINS CYSTEINE SWITCH DM00558   P22757   100-337: A184-T334, P76- P124 MATRIXINS CYSTEINE SWITCH DM00558   P08254   29-274: Q158-T334, L85- M122 MATRIX METALLOPROTEINASE PD168921: S327- BLAST_PRODOM N392 MATRIX METALLOPROTEINASE PD169970: A494- BLAST_PRODOM M568	HMMER_PFAM HMMER_PFAM BLAST_DOMO BLAST_DOMO
					MATRIX PRECURSOR METALLOPROTEASE HYDROLASE ZINC ZYMOGEN CALCIUM COLLAGEN DEGRADATION SIGNAL PD000673: F171-T266, P73-M122	BLAST_PRODOM
					Matrixins cysteine switch BL00546: F92-D121, V224-P267, G273-Y304, L313-G326, F443-Y455, F409-E428	BLIMPS_BLOCKS
					Hemopexin domain protein BL00024: M112-Y408-D419	BLIMPS_BLOCKS
					MATRIXIN SIGNATURE PR00138: M112-P125, E198-F213, V224-W252, V279-Y304, L313-G326	BLIMPS_PRINTS
					Matrixins cysteine switch cysteine_switch.prf: A95-M204	PROFILESCAN
					Neutral zinc metallopeptidases, zinc-binding region signature zinc_protease.prf: D256-E312	PROFILESCAN

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					Hemopexin domain signature hemopexin.prf: F409-R477 signal_peptide: M1-P21 Zn binding region Zinc_Protease: V279-L288 signal_cleavage: M1-P24	PROFILESCAN HMMER MOTIFS SPScan HMMER_PFM
20	4333459CD1	306	S117 S138 S2 S223 S60 S72 T110 T139 T207 T217	N108	TRYPSIN FAMILY SERINE PROTEASE trypsin:I56-I298 TRYPSIN DM00018 Q05319 543-784: I56-I302 PROTEASE SERINE PRECURSOR SIGNAL HYDROLASE ZYMOGEN GLYCOPROTEIN FAMILY MULTIGENE FACTOR PD000046: S117-I298, I56-G192 Serine proteases, trypsin family BLIMPSS_BLOCKS BL00134: C81-C97, D238-G261, P285-I298 Type I fibronectin domain BL01253: C81-A94, G154-E190, R237-C250 Kringle domain proteins BL00021: C81-I98, I165-G186, S247-F288 CHYMOTRYPSIN SERINE PROTEASE PR00722: G82-C97, P142-F156, R237-M249 Trypsin family serine protease active sites trypsin_his.prf: L73-G122 trypsin_ser.prf: R225-R271 Trypsin family serine protease active sites Trypsin_His: I92-C97 Trypsin_Ser: D238-M249 signal_cleavage: M1-S26	PROFILESCAN MOTIFS SPScan

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
21	6817347CD1	953	S102 S114 S150 S172 S369 S429 S47 S623 S794 S804 S808 S831	N95	Ubiquitin carboxyl-terminal hydrolase Family 1 UCH-1: R593-D624	HMMER_PFAM
			S856 S919 S942 T289 T42 T455 T488 T544 T567 T568 T585 T59		Ubiquitin carboxyl-terminal hydrolase Family 2 UCH-2: N875-K935	HMMER_PFAM
			T736 T777 T786 T839 Y929		UBIQUITIN CARBOXYL-TERMINAL HYDROLASES FAMILY 2 DM00659 P40818 782-1103: T777-L931, L598-H709, T713-T753, V101- L128	BLAST_DOMO
					PROTEASE UBIQUITIN HYDROLASE UBIQUITIN SPECIFIC ENZYME DEUBIQUITINATING C- TERMINAL THIOLESTERASE PROCESSING CONJUGATION PDD017412: T777-E859	BLAST_PRODOM
					Ubiquitin carboxyl-terminal hydrolase Family 2 BL00972: G594-L611, Y675-L684, I714-C728, K878-H902, K904-D925	BLIMPS_BLOCKS
					Ubiquitin carboxyl-terminal hydrolase Family 2 signature 1 Uch_2_1: G594-Q609	MOTIFS
					Ubiquitin carboxyl-terminal hydrolase Family 2 signature 2 Uch_2_2: Y879-Y896	MOTIFS

Table 4

Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Sequence Length	Selected Fragment(s)	Sequence Fragments	5' Position	3' Position
22	275791CB1	2204	1168-1197, 1503-1522, 1-281, 1716-1738	6456514H1 (COLNDTC01)	890	1510
				7246467T8 (PROSTMV01)	692	1390
				4943009F8 (BRAIFFEN05)	1319	1877
				55047202J1	1	811
23	1389845CB1	2036	1-392, 1468-1491, 1334-1400, 1974-2036	6053385H1 (BRABDIR03)	1636	2204
				FL1389845_00001	6	2036
				1389845H1 (EOSINOT01)	1	244
24	1726609CB1	2185	1-44, 1804- 2185	71762189V1	1	662
				5426388F9 (PROSTM07)	1370	1992
				71053940V1	744	1373
				71041539V1	1954	2185
				5968441H1 (BRAZNOT01)	1326	1824
				6756865U1 (SINTFER02)	642	1316
				2053131H1 (BEPINOT01)	2885	3136
				6440674H1 (BRAENOT02)	2576	3095
				95745066	1831	2254
				7191212H2 (BRATDIC01)	2295	2873
25	4503848CB1	3486	1-1330	5960039H1 (BRATNOT05)	1229	1797
				GBL.g7710158_edit	1	2015
				60200050D1	825	1146
				5969176H1 (BRAZNOT01)	1677	2104
				5649471H1 (BRAITUT23)	3157	3486
				2232143F6 (PROSNOT16)	2235	2670
				3022114H1 (PROSDINO1)	3116	3402
				60220456D1	1085	1462

Table 4 (cont.)

Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Sequence Length	Selected Fragment(s)	Sequence Fragments	5' Position	3' Position
26	5544089CB1	2847	1260-1631, 2532-2847, 408-914	55051688J1 2344450F6 (TESTTUT02) 7658834H1 (OVARNOE02) 71763578V1 6576463H1 (COLHTUS02)	810 2195 2130 1562 1079	1767 2847 2642 2266 1838
27	7474081CB1	890	1-21	g2103202 g2142177	1 397	493 890
28	5281209CB1	1577	1-629	FL5281209_97712102_000 004_g436191	1	1467
29	2256251CB1	1958	1-399, 896-935	3142983R6 (HNT2AZS07) 3220504T6 (COLMNNO03) 94264312 FL2256251_97708357_000 002_g6103629	1045 1611 400 910	1577 1576 1958 1830
30	7160544CB1	3106	1-540, 1166-1428	2256251R6 (OVARHT01) 6471337H1 (PLACFB01) 6854305H1 (BRAIFEN08) 4443368H1 (SINDNOC01) 6894004J1 (BRAITDR03) 7655990H1 (UTREDME06) 7745974H1 (ADRETTUE04) 7160544H1 (HNT2TXC01) 70490289V1 70748463V1 7745974J1 (ADRETTUE04)	408 1654 895 1332 470 324 2378 1 2636 2269 1566	1003 2316 1561 1585 1077 822 3052 427 3106 2891 2153

Table 4 (cont.)

Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Sequence Length	Selected Fragment(s)	Sequence Fragments	5' Position	3' Position	
31	7477386CB1	3567	1-971, 1953-2846, 3243-3567	GBI:g6682143_000029_ed it.20231-20345 GBI:g6682143_000023.co mp_edit.11365-11445 GBI:g6682143_000027_ed it.14110-14265 GBI:g6682143_000029_ed it.35032-35202 GBI:g6682143_000029_ed it.13651-13803 GBI:g6682143_000019_ed it.3461-3655 GBI:g6682143_000029_ed it.41513_41644 GBI:g6682143_000029_ed it.43912_44404 GBI:g6682143_000029_ed it.12846-12930 GBI:g6682143_000029_ed it.15804-15912 GBI:g6682143_000023.co mp_edit.9253-9693 GBI:g6682143_000029_ed it.27621-27746 GBI:g6682143_000029_ed it.18722-18853 GBI:g6682143_000029_ed it.17438-17581 GBI:g6682143_000029_ed it.35651-35783	1495 1 523 2272 958 679 2944 3076 874 1111 3075 3567 957 1111 82 2068 1363 1219 2608 2739	1608 81 678 2436 1110 873 3075 3567 957 1218 522 2193 1494 1362 2739	

Table 4

Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Sequence Length	Selected Fragment(s)	Sequence Fragments	5' Position	3' Position
22	275791CB1	2204	1168-1197, 1503-1522, 1-281, 1716-1738	6456514H1 (COLNDIC01) 7246467T8 (PROSTMY01) 4943009F8 (BRAIFEN05) 55047202J1 6053385H1 (BRABDIR03)	890 692 1319 1 1636	1510 1390 1877 811 2204
23	1389845CB1	2036	1-392, 1468-1491, 1334-1400, 1974-2036	FL1389845_00001 1389845H1 (EOSINOT01)	6 1	2036 244
24	1726609CB1	2185	1-44, 1804- 2185	71762189V1 5426388F9 (PROSTM07) 71053940V1 71041539V1 71041539V1 5968441H1 (BRAZNOT01) 6756865J1 (SINTFER02)	1 1370 744 1954 1326 642 2885	662 1992 1373 2185 1824 1316 3136
25	4503848CB1	3486	1-1330	2053131H1 (BEPINOT01) 6440674H1 (BRAENOT02) 95745066 7191212H2 (BRATDIC01) 5960039H1 (BRATNOT05) GBI.g7710158_edit 60200050D1 5969176H1 (BRAZNOT01) 5649471H1 (BRAITUT23) 2232143F6 (PROSNOT16) 3022114H1 (PROSDIN01) 60220456D1	1 1831 2295 1229 1 825 1677 3157 2235 3116 1085	3095 2254 2873 1797 2015 1146 2104 3486 2670 3402 1462

Table 4 (cont.)

Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Sequence Length	Selected Fragment(s)	Sequence Fragments	5' Position	3' Position
26	5544089CB1	2847	1260-1631, 2532-2847, 408-914	55051688V1 2344450F6 (TESTTUT02) 7658834H1 (OVARNOE02) 71763578V1	810 2195 2130 1562	1767 2847 2642 2266
27	7474081CB1	890	1-21	55051680J1 (COLHTU502) 55051680J1.comp	1079 1	1838 958
28	5281209CB1	1577	1-629	92103202 92142177	1 397	493 890
29	2256251CB1	1958	1-399, 896-935	FL5281209_97712102_000 004_9436191 93644494 3142983R6 (HNT2AZS07)	1 1155 1045	1467 1577 1576
30	7160544CB1	3106	1-540, 1166-1428	3220504T6 (COLNNNON03) 94264312 FL2256251_97708357_000 002_96103629 2256251R6 (OVERTUT01)	1611 400 910 408	1958 848 1830 1003
				6471337H1 (PLACFEB01) 6854305H1 (BRAIFEN08) 4443368H1 (SINDNOT01)	1654 895 1332	2316 1561 1585
				6894004J1 (BRAITDRO3) 7655990H1 (UTREDME06) 7745974H1 (ADRETUE04) 7160544H1 (HNT2ITX01)	470 324 2378 1	1077 822 3052 427
				70490289V1 70748463V1 7745974J1 (ADRETUE04)	2636 2269 1566	3106 2891 2153

Table 4 (cont.)

Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Sequence Length	Selected Fragment(s)	Sequence Fragments	5' Position	3' Position
31	7477386CB1	3567	1-971, 1953-2846, 3243-3567	GBI:g6682143_000029_ed it.20231-20345 GBI:g6682143_000023.co mp_edit.11365-11445 GBI:g6682143_000027_ed it.14110-14265 GBI:g6682143_000029_ed it.35032-35202 GBI:g6682143_000029_ed it.13651-13803 GBI:g6682143_000019_ed it.3461-3655 GBI:g6682143_000029_ed it.41513-41644 GBI:g6682143_000029_ed it.43912-44404 GBI:g6682143_000029_ed it.12846-12930 GBI:g6682143_000029_ed it.15804-15912 GBI:g6682143_000023.co mp_edit.9253-9693 GBI:g6682143_000029_ed it.27621-27746 GBI:g6682143_000029_ed it.18722-18853 GBI:g6682143_000029_ed it.17438-17581 GBI:g6682143_000029_ed it.35651-35783	1495 1 523 2272 958 679 1110 873 2944 3075 3076 874 1111 82 2068 1363 1219 2608 2739	1608 81 678 2436 1110 873 3075 3567 957 1218 522 2193 1494 1362 2739

Table 4 (cont.)

Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Sequence Length	Selected Fragment(s)	Sequence Fragments	5' Position	3' Position
				GBI:g6682143_000029_ed it.27099-27237	1759	1932
				GBI:g6682143_000029_ed it.24540-24713	1609	1758
				GBI:g6682143_000029_ed it.37355-37558	2740	2943
32	7473089CB1	2930	1-632, 1082-1138, 2453-2555, 1373-1615, 1716-1740	GBI:g7387384_000011.co mp.edit.13491-13892 GBI:g7387384_000010_ed it.2924-3211 GBI:g7387384_000010_ed it.13479-13850 GBI:g7387384_000010_ed it.11350-11535 7988641H1 (UTRSTUC01) GBI:g7387384_000010_ed it.9694-9867 GBI:g7387384_000012.co mp.edit.9639-10595 GBI:g7387384_000010_ed it.1917-2074 GBI:g7387384_000010_ed it.2514-2684 7631548J1 (BRAFTUE03) GBI:g7387384_000010_ed it.11639-11815 GBI:g7387384_edit	2529 1335 2157 1794 1064 1620 75 1032 1163 1164 1031 21 1980 1 2930	2930 1619 2528 1979 1587 1793 1031 1163 1334 619 2156 2930

Table 4 (cont.)

Polymer ID No:	Incyte Polynucleotide ID	Sequence Length	Selected Fragment(s)	Sequence Fragments	5' Position	3' Position
33	7604035CB1	4230	4185-4230, 894-2774	6254235H1 (LUNPTUT02) 6213818H1 (MUSCDIT06) 6314348H1 (NERDTDN03) 7195502H1 (LUNGFER04) 6800634J1 (COLENOR03) 8113675H1 (OSTEUNC01) 6804059H1 (COLENOR03) 7750274H1 (HEAONOE01) 3843227F6 (DENDNOT01) 7632961H1 (BLADTUE01) 55097977J1 7716357J1 (SINTFEE02)	3308 3900 3426 2758 2678 1661 478 1995 2094 1418 688 1 1340 166 2757 2650 2166 579 3156 1998 555 1 1	3897 4230 3994 3376 3323 2049 1050 2503 2707 2024 1535 678 2118 760 3441 3239 2732 1013 3699 2671 1850 473
34	3473847CB1	3699	1-2631	71906145V1 7101935F8 (BRAWTDR02) 70857826V1 70855756V1 820867R1 (KERANOT02) 8055446J1 (ESOGTUE01) 70857738V1 70858612V1 GNN.g7208751_000002_00 2.edit 7101935R8 (BRAWTDR02)		

Table 4 (cont.)

Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Sequence Length	Selected Fragment (s)	Sequence Fragments	5' Position	3' Position
35	3750004CB1	2410	1-264, 2116-2410, 1057-1167, 1590-1649	g7712021.edit 7680089J1 (BRAFTUE01) 6804411H1 (COLENOR03) 71909368V1 g1187194 g2241985 6314962H1 (NERDTDN03) 6823371J1 (SINTNOR01) 7655009J1 (UTREDME06) g1272147 71620969V1	1 1327 1088 536 1655 706 973 65 1407 1754 1 1	246 1911 1618 968 2127 1144 1135 855 1990 2410 549 1340 2514 2755 1500 139 819 2263 2120
36	4904126CB1	549				
37	71268415CB1	2755	1-1097, 2326-2755	7715927J1 (SINTFEE02) 7372052H2 (BRAIFFE04) g6651070_CD 7723192J2 (THYRDIE01) GBI:g7709257_000011.ed it	590 2044 102 905 1  8037549H1 (SMCRUNE01) 7720289J1 (THYRDIE01) 8037549J1 (SMCRUNE01)	1990 2410 549 1340 2514 2755 1500 139 819 2263 2120

Table 4 (cont.)

Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Sequence Length	Selected Fragment(s)	Sequence Fragments	5' Position	3' Position
38	7473301CB1	2553	1-2394	GBI.g7272157_000017.ed it	2001	2553
				71704195V1	1713	2016
				5544473H1 (TESTNOC01)	622	680
				GNN.g7272157_000017_00 2.edit	1688	2382
				5544473T8 (TESTNOC01)	2246	2550
				71703469V1	1163	1746
				GNN.g8571511_000004_00 2.edit	981	1468
				GNN.g6624046_000008_00 4	1	1111
39	7473308CB1	1041	826-1041, 1-299	GNN.g1552511_035	1	1041
40	7478021CB1	1707	1-1188	g8176728_edit 97684439_edit 97684439_edit_2	979	1083
				97684439	1	978
				1084	1707	
41	4333459CB1	1262	1-1262	71571956V1 5634861R8 (PLACFER01)	704	1262
				71571988V1	1	266
				256	937	
				71573159V1	247	928
				55022864H1	2314	3067
				55022792H2	1392	2091
				55022814H1	2080	2886
				55022795J2	2044	2726
				GNN.g7417337_004.edit	1	3067

Table 5

Polynucleotide SEQ ID NO:	Incyte Project ID	Representative Library
22	275791CB1	TESTNOT03
23	1389845CB1	EOSITXT01
24	1726609CB1	BRALTUT02
25	4503848CB1	PROSNOT16
26	5544089CB1	BRAIFFEC01
28	5281209CB1	HNT2AZS07
29	2256251CB1	OVARTUT01
30	7160544CB1	BRAFNOT02
32	7473089CB1	UTRSTTUC01
33	7604035CB1	FLACNOR01
34	3473847CB1	KERANOT02
35	3750004CB1	BRAFTUE01
36	4904126CB1	TLYMNNOT08
37	71268415CB1	THYRDIE01
38	7473301CB1	TESTNOC01
41	4333459CB1	KIDCTMT01
42	6817347CB1	ADRETRUR01

Table 6

Library	Vector	Library Description
ADRETUR01	PCDNA2.1	This random primed library was constructed using RNA isolated from left upper pole, adrenal gland tumor tissue removed from a 52-year-old Caucasian male during nephroureterectomy and local destruction of renal lesion. Pathology indicated grade 3 adrenal cortical carcinoma forming a mass that infiltrated almost the whole adrenal parenchyma and extended to adjacent adipose tissue. A metastatic tumor nodule was identified in the hilar region. The renal vein was infiltrated by tumor and the neoplastic process was present at the resection margin of the renal vein. Fragments of adrenal cortical carcinoma and thrombus were found in the inferior vena cava. Patient history included abnormal weight loss. Family history included skin cancer, type I diabetes, and neurotic depression.
BRAFNOT02	pINCY	Library was constructed using RNA isolated from superior frontal cortex tissue removed from a 35-year-old Caucasian male who died from cardiac failure. Pathology indicated moderate leptomeningeal fibrosis and multiple microinfarctions of the cerebral neocortex. Microscopically, the cerebral hemisphere revealed moderate fibrosis of the leptomeninges with focal calcifications. There was evidence of shrunken and slightly eosinophilic pyramidal neurons throughout the cerebral hemispheres. In addition, scattered throughout the cerebral cortex, there were multiple small microscopic areas of cavitation with surrounding gliosis. Patient history included dilated cardiomyopathy, congestive heart failure, cardiomegaly, and an enlarged spleen and liver.
BRAFTUE01	PCDNA2.1	This 5' biased random primed library was constructed using RNA isolated from brain tumor tissue removed from the frontal lobe of a 58-year-old Caucasian male during excision of a cerebral meningeal lesion. Pathology indicated a grade 2 metastatic hemangioma. The patient presented with migraine headache. The patient developed a cerebral hemorrhage and pulmonary edema, and died during this hospitalization. Patient history included a grade 2 renal cell carcinoma, insomnia, and chronic airway obstruction. Previous surgeries included a nephroureterectomy. Patient medications included Decadron and Dilantin. Family history included a malignant neoplasm of the kidney in the father.
BRAIFEC01	pINCY	This large size-fractionated library was constructed using RNA isolated from brain tissue removed from a Caucasian male fetus who was stillborn with a hypoplastic left heart at 23 weeks' gestation.

Table 6 (cont.)

Library	Vector	Library Description
BRAITUT02	PSPORT1	Library was constructed using RNA isolated from brain tumor tissue removed from the frontal lobe of a 58-year-old Caucasian male during excision of a cerebral meningeal lesion. Pathology indicated a grade 2 metastatic hypernephroma. Patient history included a grade 2 renal cell carcinoma, insomnia, and chronic airway obstruction. Family history included a malignant neoplasm of the kidney.
EOSITXT01	DINCY	Library was constructed using RNA isolated from eosinophils stimulated with IL-5.
HNT2AZS07	PSPORT1	This subtracted library was constructed from RNA isolated from an HNT2 cell line (derived from a human teratocarcinoma that exhibited properties characteristic of a committed neuronal precursor) treated for three days with 0.35 micromolar AZ. The hybridization probe for subtraction was derived from a similarly constructed library from untreated HNT2 cells. 3.08M clones from the AZ-treated library were subjected to three rounds of subtractive hybridization with 3.04M clones from the untreated library. Subtractive hybridization conditions were based on the methodologies of Swaroop et al. (NAR (1991) 19:1954) and Bonaldo et al. (Genome Research (1996) 6:791).
KERANOT02	PSPORT1	Library was constructed using RNA isolated from epidermal breast keratinocytes (NHEK). NHEK (Clontech #CC-2501) is human breast keratinocyte cell line derived from a 30-year-old black female during breast-reduction surgery.
KIDCTMT01	PINCY	Library was constructed using RNA isolated from kidney cortex tissue removed from a 65-year-old male during nephroureterectomy. Pathology for the associated tumor tissue indicated grade 3 renal cell carcinoma within the mid-portion of the kidney and the renal capsule.
OVARTUT01	PSPORT1	Library was constructed using RNA isolated from ovarian tumor tissue removed from a 43-year-old Caucasian female during removal of the Fallopian tubes and ovaries. Pathology indicated grade 2 mucinous cystadenocarcinoma involving the entire left ovary. Patient history included mitral valve disorder, pneumonia, and viral hepatitis. Family history included atherosclerotic coronary artery disease, pancreatic cancer, stress reaction, cerebrovascular disease, breast cancer, and uterine cancer.

Table 6 (cont.)

Library	Vector	Library Description
PLACNOR01	PCDNA2.1	This random primed library was constructed using pooled cDNA from two different donors. cDNA was generated using mRNA isolated from placental tissue removed from a Caucasian fetus (donor A), who died after 16 weeks' gestation from fetal demise and hydrocephalus and from placental tissue removed from a Caucasian male fetus (donor B), who died after 18 weeks' gestation from fetal demise. Patient history for donor A included umbilical cord wrapped around the head (3 times) and the shoulders (1 time). Serology was positive for anti-CMV and remaining serologies were negative. Family history included multiple pregnancies and live births, and an abortion in the mother. Serology was negative for donor B.
PROSNOT16	PINCY	Library was constructed using RNA isolated from diseased prostate tissue removed from a 68-year-old Caucasian male during a radical prostatectomy. Pathology indicated adenofibromatous hyperplasia. Pathology for the associated tumor tissue indicated an adenocarcinoma (Gleason grade 3+4). The patient presented with elevated prostate specific antigen (PSA). During this hospitalization, the patient was diagnosed with myasthenia gravis. Patient history included osteoarthritis, and type II diabetes. Family history included benign hypertension, acute myocardial infarction, hyperlipidemia, and arteriosclerotic coronary artery disease.
TESTNOC01	PBLUESCRIPT	This large size fractionated library was constructed using RNA isolated from testicular tissue removed from a pool of eleven, 10 to 61-year-old Caucasian males.
TESTNOT03	PBLUESCRIPT	Library was constructed using RNA isolated from testicular tissue removed from a 37-year-old Caucasian male, who died from liver disease. Patient history included cirrhosis, jaundice, and liver failure.
THYRDIE01	PCDNA2.1	This 5' biased random primed library was constructed using RNA isolated from diseased thyroid tissue removed from a 22-year-old Caucasian female during closed thyroid biopsy, partial thyroidectomy, and regional lymph node excision. Pathology indicated adenomatous hyperplasia. The patient presented with malignant neoplasm of the thyroid. Patient history included normal delivery, alcohol abuse, and tobacco abuse. Previous surgeries included myringotomy. Patient medications included an unspecified type of birth control pills. Family history included hyperlipidemia and depressive disorder in the mother; and benign hypertension, congestive heart failure, and chronic leukemia in the grandparent(s).

Table 6 (cont.)

Library	Vector	Library Description
TLYMNCT08	PINCY	The library was constructed using RNA isolated from anergic allogenic T-lymphocyte tissue removed from an adult (40-50-year-old) Caucasian male. The cells were incubated for 3 days in the presence of 1 microgram/ml OKT3 mAb and 5% human serum.
UTRSTUC01	PSPORT1	This large size fractionated library was constructed using pooled cDNA from two donors. cDNA was generated using mRNA isolated from uterus tumor tissue removed from a 37-year-old Black female (donor A) during myomectomy, dilation and curettage, right fimbrial region biopsy, and incidental appendectomy; and from endometrial tumor tissue removed from a 49-year-old Caucasian female (donor B) during vaginal hysterectomy and bilateral salpingo-oophorectomy. For donor A, pathology indicated multiple uterine leiomyomata. A fimbrial cyst was identified. The endometrium was in secretory phase with hormonal effect. The patient presented with deficiency anemia, an umbilical hernia, and premenopausal menorrhagia. Patient history included premenopausal menorrhagia and sarcoidosis of the lung. Previous surgeries included hysteroscopy, dilation and curettage, and endoscopic lung biopsy. Patient medications included Chromagen and Claritin. For donor B, pathology indicated grade 3 adenosquamous carcinoma forming a mass within the uterine fundus and involving the anterior uterine wall, as well as focally involving an adjacent endometrial polyp. The tumor invaded to a maximum depth of 7mm (uninvolved wall thickness, 2.2cm). The adjacent endometrium was inactive. Paraffin section immunostains for estrogen receptors and progesterone receptors were positive. Patient history included malignant breast neoplasm. Previous surgeries included unilateral extended simple mastectomy and bilateral tubal destruction. Patient medications included Megace and CAF (Cyclophosphamide, Adriamycin, Fluoroacil).

Table 7

Program	Description	Reference	Parameter Threshold
ABI FACTURA	A program that removes vector sequences and masks ambiguous bases in nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
ABI/PARACEL FDF	A Fast Data Finder useful in comparing and annotating amino acid or nucleic acid sequences.	Applied Biosystems, Foster City, CA; Paracel Inc., Pasadena, CA.	Mismatch <50%
ABI AutoAssembler	A program that assembles nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
BLAST	A Basic Local Alignment Search Tool useful in sequence similarity search for amino acid and nucleic acid sequences. BLAST includes five functions: blastp, blastn, blastn, tblastn, and tblastx.	Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410; Altschul, S.F. et al. (1997) Nucleic Acids Res. 25:3389-3402.	<i>ESTs</i> : Probability value= 1.0E-8 or less <i>Full Length sequences</i> : Probability value= 1.0E-10 or less
FAS TA	A Pearson and Lipman algorithm that searches for similarity between a query sequence and a group of sequences of the same type. FASTA comprises at least five functions: fasta, tfasta, fastx, tfastx, and ssearch.	Pearson, W.R. and D.J. Lipman (1988) Proc. Natl. Acad. Sci. USA 85:2444-2448; Pearson, W.R. (1990) Methods Enzymol. 183:63-98; and Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489.	<i>ESTs</i> : fasta E value=1.0E-6 <i>Assembled ESTs</i> : fasta Identity= 95% or greater and Match length=200 bases or greater; fastx E value=1.0E-8 or less <i>Full Length sequences</i> : fastx score=100 or greater
BLIMPS	A BLocks IMProved Searcher that matches a sequence against those in BLOCKS, PRINTS, DOMO, PRODOM, and PFAM databases to search for gene families, sequence homology, and structural fingerprint regions.	Henikoff, S and J.G. Henikoff (1991) Nucleic Acids Res. 19:6565-6572; Henikoff, J.G. and S. Henikoff (1996) Methods Enzymol. 266:88-105; and Attwood, T.K. et al. (1997) J. Chem. Inf. Comput. Sci. 37:417-424.	Probability value= 1.0E-3 or less
HMMER	An algorithm for searching a query sequence against hidden Markov model (HMM)-based databases of protein family consensus sequences, such as PFAM.	Krogh, A. et al. (1994) J. Mol. Biol. 235:1501-1531; Sonhammer, E.L.L. et al. (1988) Nucleic Acids Res. 26:320-322; Durbin, R. et al. (1998) Our World View, in a Nutshell, Cambridge Univ. Press, pp. 1-350.	<i>PFAM hits</i> : Probability value= 1.0E-3 or less <i>Signal peptide hits</i> : Score= 0 or greater

**Table 7 (cont.)**

Program	Description	Reference	Parameter Threshold
ProfileScan	An algorithm that searches for structural and sequence motifs in protein sequences that match sequence patterns defined in Prosite.	Gribskov, M. et al. (1988) CABIOS 4:61-66; Gribskov, M. et al. (1989) Methods Enzymol. 183:146-159; Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221.	Normalized quality score $\geq$ GCGR-specified "HIGH" value for that particular Prosite motif. Generally, score=1.4-2.1.
Phred	A base-calling algorithm that examines automated sequencer traces with high sensitivity and probability.	Ewing, B. et al. (1998) Genome Res. 8:175-185; Ewing, B. and P. Green (1998) Genome Res. 8:186-194.	
Phrap	A Phil's Revised Assembly Program including SWAT and CrossMatch, programs based on efficient implementation of the Smith-Waterman algorithm, useful in searching sequence homology and assembling DNA sequences.	Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:462-489; Smith, T.F. and M.S. Waterman (1981) J. Mol. Biol. 147:195-197; and Green, P., University of Washington, Seattle, WA.	Score=120 or greater; Match length= 56 or greater
Consed	A graphical tool for viewing and editing Phrap assemblies.	Gordon, D. et al. (1998) Genome Res. 8:195-202.	
SPPScan	A weight matrix analysis program that scans protein sequences for the presence of secretory signal peptides.	Nielson, H. et al. (1997) Protein Engineering 10:1-6; Claverie, J.M. and S. Audic (1997) CABIOS 12:431-439.	Score=3.5 or greater
TMAP	A program that uses weight matrices to delineate transmembrane segments on protein sequences and determine orientation.	Person, B. and P. Argos (1994) J. Mol. Biol. 237:182-192; Person, B. and P. Argos (1996) Protein Sci. 5:363-371.	
TMHMMER	A program that uses a hidden Markov model (HMM) to delineate transmembrane segments on protein sequences and determine orientation.	Somhammer, E.L. et al. (1998) Proc. Sixth Int'l. Conf. on Intelligent Systems for Mol. Biol., Glasgow et al., eds., The Am. Assoc. for Artificial Intelligence Press, Menlo Park, CA, pp. 175-182.	
Motifs	A program that searches amino acid sequences for patterns that matched those defined in Prosite.	Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221; Wisconsin Package Program Manual, version 9, page M51-59, Genetics Computer Group, Madison, WI.	

What is claimed is:

1. An isolated polypeptide selected from the group consisting of:
  - a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-21,
    - b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-21,
      - c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, and
      - d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-21.
  2. An isolated polypeptide of claim 1 selected from the group consisting of SEQ ID NO:1-21.
  3. An isolated polynucleotide encoding a polypeptide of claim 1.
  4. An isolated polynucleotide encoding a polypeptide of claim 2.
  5. An isolated polynucleotide of claim 4 selected from the group consisting of SEQ ID NO:22-42.
  6. A recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide of claim 3.
  7. A cell transformed with a recombinant polynucleotide of claim 6.
  8. A transgenic organism comprising a recombinant polynucleotide of claim 6.
  9. A method for producing a polypeptide of claim 1, the method comprising:
    - a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide, and said recombinant polynucleotide comprises a promoter sequence operably linked to a polynucleotide encoding the polypeptide of claim 1, and
    - b) recovering the polypeptide so expressed.

10. An isolated antibody which specifically binds to a polypeptide of claim 1.
11. An isolated polynucleotide selected from the group consisting of:
  - a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:22-42,
  - b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:22-42,
  - c) a polynucleotide complementary to a polynucleotide of a),
  - d) a polynucleotide complementary to a polynucleotide of b), and
  - e) an RNA equivalent of a)-d).
12. An isolated polynucleotide comprising at least 60 contiguous nucleotides of a polynucleotide of claim 11.
- 15 13. A method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 11, the method comprising:
  - a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and
  - b) detecting the presence or absence of said hybridization complex, and, optionally, if present, the amount thereof.
- 25 14. A method of claim 13, wherein the probe comprises at least 60 contiguous nucleotides.
15. A method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 11, the method comprising:
  - a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and
  - b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof.
- 30 16. A composition comprising a polypeptide of claim 1 and a pharmaceutically acceptable excipient.

17. A composition of claim 16, wherein the polypeptide has an amino acid sequence selected from the group consisting of SEQ ID NO:1-21.

18. A method for treating a disease or condition associated with decreased expression of 5 functional PRTS, comprising administering to a patient in need of such treatment the composition of claim 16.

19. A method for screening a compound for effectiveness as an agonist of a polypeptide of claim 1, the method comprising:

10        a) exposing a sample comprising a polypeptide of claim 1 to a compound, and  
            b) detecting agonist activity in the sample.

20. A composition comprising an agonist compound identified by a method of claim 19 and a pharmaceutically acceptable excipient.

15        21. A method for treating a disease or condition associated with decreased expression of functional PRTS, comprising administering to a patient in need of such treatment a composition of claim 20.

20        22. A method for screening a compound for effectiveness as an antagonist of a polypeptide of claim 1; the method comprising:

            a) exposing a sample comprising a polypeptide of claim 1 to a compound, and  
            b) detecting antagonist activity in the sample.

25        23. A composition comprising an antagonist compound identified by a method of claim 22 and a pharmaceutically acceptable excipient.

24. A method for treating a disease or condition associated with overexpression of functional PRTS, comprising administering to a patient in need of such treatment a composition of claim 23.

30        25. A method of screening for a compound that specifically binds to the polypeptide of claim 1, said method comprising the steps of:  
            a) combining the polypeptide of claim 1 with at least one test compound under suitable conditions, and

b) detecting binding of the polypeptide of claim 1 to the test compound, thereby identifying a compound that specifically binds to the polypeptide of claim 1.

26. A method of screening for a compound that modulates the activity of the polypeptide of  
5 claim 1, said method comprising:

- a) combining the polypeptide of claim 1 with at least one test compound under conditions permissive for the activity of the polypeptide of claim 1,
- b) assessing the activity of the polypeptide of claim 1 in the presence of the test compound, and
- c) comparing the activity of the polypeptide of claim 1 in the presence of the test compound  
10 with the activity of the polypeptide of claim 1 in the absence of the test compound, wherein a change in the activity of the polypeptide of claim 1 in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide of claim 1.

27. A method for screening a compound for effectiveness in altering expression of a target  
15 polynucleotide, wherein said target polynucleotide comprises a sequence of claim 5, the method comprising:

- a) exposing a sample comprising the target polynucleotide to a compound, under conditions suitable for the expression of the target polynucleotide,
- b) detecting altered expression of the target polynucleotide, and
- c) comparing the expression of the target polynucleotide in the presence of varying amounts of  
20 the compound and in the absence of the compound.

28. A method for assessing toxicity of a test compound, said method comprising:  
a) treating a biological sample containing nucleic acids with the test compound;

- b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide of claim 11 under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide comprising a polynucleotide sequence of a polynucleotide of claim 11 or fragment thereof;
- c) quantifying the amount of hybridization complex; and
- d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

29. A diagnostic test for a condition or disease associated with the expression of PRTS in a biological sample comprising the steps of:

- a) combining the biological sample with an antibody of claim 10, under conditions suitable  
5 for the antibody to bind the polypeptide and form an antibody:polypeptide complex; and
- b) detecting the complex, wherein the presence of the complex correlates with the presence  
of the polypeptide in the biological sample.

30. The antibody of claim 10, wherein the antibody is:

- 10 a) a chimeric antibody,
- b) a single chain antibody,
- c) a Fab fragment,
- d) a F(ab')<sub>2</sub> fragment, or
- e) a humanized antibody.

15

31. A composition comprising an antibody of claim 10 and an acceptable excipient.

32. A method of diagnosing a condition or disease associated with the expression of PRTS in a subject, comprising administering to said subject an effective amount of the composition of claim

20 31.

33. A composition of claim 31, wherein the antibody is labeled.

34. A method of diagnosing a condition or disease associated with the expression of PRTS in a subject, comprising administering to said subject an effective amount of the composition of claim  
25 33.

35. A method of preparing a polyclonal antibody with the specificity of the antibody of claim 10 comprising:

- 30 a) immunizing an animal with a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, or an immunogenic fragment thereof, under conditions to elicit an antibody response;
- b) isolating antibodies from said animal; and

c) screening the isolated antibodies with the polypeptide, thereby identifying a polyclonal antibody which binds specifically to a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-21.

5 36. An antibody produced by a method of claim 35.

37. A composition comprising the antibody of claim 36 and a suitable carrier.

38. A method of making a monoclonal antibody with the specificity of the antibody of claim  
10 10 comprising:

a) immunizing an animal with a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, or an immunogenic fragment thereof, under conditions to elicit an antibody response;

15 b) isolating antibody producing cells from the animal;

c) fusing the antibody producing cells with immortalized cells to form monoclonal antibody-producing hybridoma cells;

d) culturing the hybridoma cells; and

e) isolating from the culture monoclonal antibody which binds specifically to a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-21.

20

39. A monoclonal antibody produced by a method of claim 38.

40. A composition comprising the antibody of claim 39 and a suitable carrier.

25 41. The antibody of claim 10, wherein the antibody is produced by screening a Fab expression library.

42. The antibody of claim 10, wherein the antibody is produced by screening a recombinant immunoglobulin library.

30

43. A method for detecting a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-21 in a sample, comprising the steps of:

a) incubating the antibody of claim 10 with a sample under conditions to allow specific binding of the antibody and the polypeptide; and

b) detecting specific binding, wherein specific binding indicates the presence of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-21 in the sample.

5 44. A method of purifying a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-21 from a sample, the method comprising:

a) incubating the antibody of claim 10 with a sample under conditions to allow specific binding of the antibody and the polypeptide; and

10 b) separating the antibody from the sample and obtaining the purified polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-21.

45. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:1.

15 46. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:2.

47. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:3.

48. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:4.

20 49. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:5.

50. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:6.

25 51. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:7.

52. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:8.

53. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:9.

30 54. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:10.

55. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:11.

56. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:12.
57. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:13.
58. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:14.
59. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:15.
60. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:16.
- 10 61. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:17.
62. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:18.
- 15 63. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:19.
64. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:20.
65. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:21.
- 20 66. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:22.
67. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:23.
- 25 68. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:24.
- 30 69. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:25.
70. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:26.

71. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:27.

72. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID 5 NO:28.

73. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:29.

10 74. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:30.

75. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:31.

15 76. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:32.

77. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID 20 NO:33.

78. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:34.

25 79. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:35.

80. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:36.

30 81. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:37.

82. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:38.

83. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID  
5 NO:39.

84. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:40.

10 85. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:41.

86. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:42.

<110> INCYTE GENOMICS, INC.  
 YUE, Henry  
 ELLIOTT, Vicki  
 GANDHI, Ameena R.  
 LAL, Preeti  
 AU-YOUNG, Janice  
 TRIBOULEY, Catherine M.  
 DELEGEANE, Angelo M.  
 BAUGHN, Mariah R.  
 NGUYEN, Danniel B.  
 LEE, Ernestine A.  
 HAFALIA, April  
 KHAN, Farrah A.  
 WALIA, Narinder K.  
 YAO, Monique G.  
 LU, Dyung Aina M.  
 PATTERSON, Chandra  
 TANG, Y. Tom  
 WALSH, Roderick T.  
 AZIMZAI, Yalda  
 LU, Yan  
 RAMKUMAR, Jayalaximi  
 XU, Yuming  
 REDDY, Roopa  
 DAS, Depopriya  
 KEARNEY, Liam  
 KALLICK, Deborah A.

&lt;120&gt; Proteases

&lt;130&gt; PI-0123 PCT

<140> To Be Assigned  
 <141> Herewith

<150> 60/212,336; 60/213,955; 60/215,396; 60/216,821; 60/218,946  
 <151> 2000-06-16; 2000-06-22; 2000-06-29; 2000-07-07; 2000-07-14

&lt;160&gt; 42

&lt;170&gt; PERL Program

<210> 1  
 <211> 232  
 <212> PRT  
 <213> Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 275791CD1

&lt;400&gt; 1

Met	Pro	Glu	Asn	Pro	Asp	Thr	Met	Glu	Thr	Glu	Lys	Pro	Lys	Thr
1							5		10					15
Ile	Thr	Glu	Leu	Asp	Pro	Ala	Ser	Phe	Thr	Glu	Ile	Thr	Lys	Asp
									25					30
Cys	Asp	Glu	Asn	Lys	Glu	Asn	Lys	Thr	Pro	Glu	Gly	Ser	Gln	Gly
								35		40				45
Glu	Val	Asp	Trp	Leu	Gln	Gln	Tyr	Asp	Met	Glu	Arg	Glu	Arg	Glu
								50		55				60
Glu	Gln	Glu	Leu	Gln	Gln	Ala	Leu	Ala	Gln	Ser	Leu	Gln	Glu	Gln
								65		70				75
Glu	Ala	Trp	Glu	Gln	Lys	Glu	Asp	Asp	Asp	Leu	Lys	Arg	Ala	Thr
								80		85				90
Glu	Leu	Ser	Leu	Gln	Glu	Phe	Asn	Asn	Ser	Phe	Val	Asp	Ala	Leu
								95		100				105
Gly	Ser	Asp	Glu	Asp	Ser	Gly	Asn	Glu	Asp	Val	Phe	Asp	Met	Glu
								110		115				120

Tyr Thr Glu Ala Glu Ala Glu Glu Leu Lys Arg Asn Ala Glu Thr  
 125 130 135  
 Gly Asn Leu Pro His Ser Tyr Arg Leu Ile Ser Val Val Ser His  
 140 145 150  
 Ile Gly Ser Thr Ser Ser Ser Gly His Tyr Ile Ser Asp Val Tyr  
 155 160 165  
 Asp Ile Lys Lys Gln Ala Trp Phe Thr Tyr Asn Asp Leu Glu Val  
 170 175 180  
 Ser Lys Ile Gln Glu Ala Ala Val Gln Ser Asp Arg Asp Arg Ser  
 185 190 195  
 Gly Tyr Ile Phe Phe Tyr Met His Lys Glu Ile Phe Asp Glu Leu  
 200 205 210  
 Leu Glu Thr Glu Lys Asn Ser Gln Ser Leu Ser Thr Glu Val Gly  
 215 220 225  
 Lys Thr Thr Arg Gln Ala Ser  
 230

&lt;210&gt; 2

&lt;211&gt; 365

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 1389845CD1

&lt;400&gt; 2

Met Pro Lys Tyr Leu Gly Gly Cys Cys Ile Pro Gly Pro Trp  
 1 5 10 15  
 Ala Glu Arg Arg Val Tyr Ser Leu Gly His Gln Asp Lys Ser Arg  
 20 25 30  
 Thr His Gln Glu Leu Arg Thr Asp Arg Arg Thr Thr Glu Gly Val  
 35 40 45  
 Thr Gly Trp Cys Glu Asp Trp Cys Pro Trp Ala Arg Thr Leu Leu  
 50 55 60  
 Ser Ser Pro Cys Trp Leu Gln Thr Arg Val Gln Ala Leu Gly Ser  
 65 70 75  
 Ala Thr Leu Thr Gln Pro Ser Leu Glu Asp Arg Met Arg Gly Val  
 80 85 90  
 Ser Cys Leu Gln Val Leu Leu Leu Val Leu Gly Ala Ala Gly  
 95 100 105  
 Thr Gln Gly Arg Lys Ser Ala Ala Cys Gly Gln Pro Arg Met Ser  
 110 115 120  
 Ser Arg Ile Val Gly Gly Arg Asp Gly Arg Asp Gly Glu Trp Pro  
 125 130 135  
 Trp Gln Ala Ser Ile Gln His Arg Gly Ala His Val Cys Gly Gly  
 140 145 150  
 Ser Leu Ile Ala Pro Gln Trp Val Leu Thr Ala Ala His Cys Phe  
 155 160 165  
 Pro Arg Arg Ala Leu Pro Ala Glu Tyr Arg Val Arg Leu Gly Ala  
 170 175 180  
 Leu Arg Leu Gly Ser Thr Ser Pro Arg Thr Leu Ser Val Pro Val  
 185 190 195  
 Arg Arg Val Leu Leu Pro Pro Asp Tyr Ser Glu Asp Gly Ala Arg  
 200 205 210  
 Gly Asp Leu Ala Leu Leu Gln Leu Arg Arg Pro Val Pro Leu Ser  
 215 220 225  
 Ala Arg Val Gln Pro Val Cys Leu Pro Val Pro Gly Ala Arg Pro  
 230 235 240  
 Pro Pro Gly Thr Pro Cys Arg Val Thr Gly Trp Gly Ser Leu Arg  
 245 250 255  
 Pro Gly Val Pro Leu Pro Glu Trp Arg Pro Leu Gln Gly Val Arg  
 260 265 270  
 Val Pro Leu Leu Asp Ser Arg Thr Cys Asp Gly Leu Tyr His Val  
 275 280 285  
 Gly Ala Asp Val Pro Gln Ala Glu Arg Ile Val Leu Pro Gly Ser  
 290 295 300

Leu Cys Ala Gly Tyr Pro Gln Gly His Lys Asp Ala Cys Gln Val  
                   305                  310                  315  
 Cys Thr Gln Pro Pro Gln Pro Pro Glu Ser Pro Pro Cys Ala Gln  
                   320                  325                  330  
 His Pro Pro Ser Leu Asn Ser Arg Thr Gln Asp Ile Pro Thr Gln  
                   335                  340                  345  
 Ala Gln Asp Pro Gly Leu Gln Pro Arg Gly Thr Thr Pro Gly Val  
                   350                  355                  360  
 Trp Asn Pro Glu Asn  
                   365

<210> 3  
 <211> 416  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 1726609CD1

<400> 3  
 Met Trp Gly Arg Tyr Asp Ile Val Phe Leu Pro Pro Ser Phe Pro.  
   1              5              10              15  
 Ile Val Ala Met Glu Asn Pro Cys Leu Thr Phe Ile Ile Ser Ser  
   20              25              30  
 Ile Leu Glu Ser Asp Glu Phe Leu Val Ile Asp Val Ile His Glu  
   35              40              45  
 Val Ala His Ser Trp Phe Gly Asn Ala Val Thr Asn Ala Thr Trp  
   50              55              60  
 Glu Glu Met Trp Leu Ser Glu Gly Leu Ala Thr Tyr Ala Gln Arg  
   65              70              75  
 Arg Ile Thr Thr Glu Thr Tyr Gly Ala Ala Phe Thr Cys Leu Glu  
   80              85              90  
 Thr Ala Phe Arg Leu Asp Ala Leu His Arg Gln Met Lys Leu Leu  
   95              100            105  
 Gly Glu Asp Ser Pro Val Ser Lys Leu Gln Val Lys Leu Glu Pro  
   110            115            120  
 Gly Val Asn Pro Ser His Leu Met Asn Leu Phe Thr Tyr Glu Lys  
   125            130            135  
 Gly Tyr Cys Phe Val Tyr Tyr Leu Ser Gln Leu Cys Gly Asp Pro  
   140            145            150  
 Gln Arg Phe Asp Asp Phe Leu Arg Ala Tyr Val Glu Lys Tyr Lys  
   155            160            165  
 Phe Thr Ser Val Val Ala Gln Asp Leu Leu Asp Ser Phe Leu Ser  
   170            175            180  
 Phe Phe Pro Glu Leu Lys Glu Gln Ser Val Asp Cys Arg Ala Gly  
   185            190            195  
 Leu Glu Phe Glu Arg Trp Leu Asn Ala Thr Gly Pro Pro Leu Ala  
   200            205            210  
 Glu Pro Asp Leu Ser Gln Gly Ser Ser Leu Thr Arg Pro Val Glu  
   215            220            225  
 Ala Leu Phe Gln Leu Trp Thr Ala Glu Pro Leu Asp Gln Ala Ala  
   230            235            240  
 Ala Ser Ala Ser Ala Ile Asp Ile Ser Lys Trp Arg Thr Phe Gln  
   245            250            255  
 Thr Ala Leu Phe Leu Asp Arg Leu Leu Asp Gly Ser Pro Leu Pro  
   260            265            270  
 Gln Glu Val Val Met Ser Leu Ser Lys Cys Tyr Ser Ser Leu Leu  
   275            280            285  
 Asp Ser Met Asn Ala Glu Ile Arg Ile Arg Trp Leu Gln Ile Val  
   290            295            300  
 Val Arg Asn Asp Tyr Tyr Pro Asp Leu His Arg Val Arg Arg Phe  
   305            310            315  
 Leu Glu Ser Gln Met Ser Arg Met Tyr Thr Ile Pro Leu Tyr Glu  
   320            325            330  
 Asp Leu Cys Thr Gly Ala Leu Lys Ser Phe Ala Leu Glu Val Phe  
   335            340            345

Tyr Gln Thr Gln Gly Arg Leu His Pro Asn Leu Arg Arg Ala Ile  
 350 355 360  
 Gln Gln Ile Leu Ser Gln Gly Leu Gly Ser Ser Thr Glu Pro Ala  
 365 370 375  
 Ser Glu Pro Ser Thr Glu Leu Gly Lys Ala Glu Ala Asp Thr Asp  
 380 385 390  
 Ser Asp Ala Gln Ala Leu Leu Leu Gly Asp Glu Ala Pro Ser Ser  
 395 400 405  
 Ala Ile Ser Leu Arg Asp Val Asn Val Ser Ala  
 410 415

&lt;210&gt; 4

&lt;211&gt; 714

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 4503848CD1

&lt;400&gt; 4

Met His Ile His Met Leu Thr Leu Asp Gln Gln Lys Ser Leu Ile  
 1 5 10 15  
 Leu Ile Leu Phe Leu Ile Leu Phe Arg Val Gly Gly Ser Arg Ile  
 20 25 30  
 Leu Leu Arg Met Thr Leu Gly Arg Glu Val Met Ser Pro Leu Gln  
 35 40 45  
 Ala Met Ser Ser Tyr Thr Val Ala Gly Arg Asn Val Leu Arg Trp  
 50 55 60  
 Asp Leu Ser Pro Glu Gln Ile Lys Thr Arg Thr Glu Glu Leu Ile  
 65 70 75  
 Val Gln Thr Lys Gln Val Tyr Asp Ala Val Gly Met Leu Gly Ile  
 80 85 90  
 Glu Glu Val Thr Tyr Glu Asn Cys Leu Gln Ala Leu Ala Asp Val  
 95 100 105  
 Glu Val Lys Tyr Ile Val Glu Arg Thr Met Leu Asp Phe Pro Gln  
 110 115 120  
 His Val Ser Ser Asp Lys Glu Val Arg Ala Ala Ser Thr Glu Ala  
 125 130 135  
 Asp Lys Arg Leu Ser Arg Phe Asp Ile Glu Met Ser Met Arg Gly  
 140 145 150  
 Asp Ile Phe Glu Arg Ile Val His Leu Gln Glu Thr Cys Asp Leu  
 155 160 165  
 Gly Lys Ile Lys Pro Glu Ala Arg Arg Tyr Leu Glu Lys Ser Ile  
 170 175 180  
 Lys Met Gly Lys Arg Asn Gly Leu His Leu Pro Glu Gln Val Gln  
 185 190 195  
 Asn Glu Ile Lys Ser Met Lys Lys Arg Met Ser Glu Leu Cys Ile  
 200 205 210  
 Asp Phe Asn Lys Asn Leu Asn Glu Asp Asp Thr Phe Leu Val Phe  
 215 220 225  
 Ser Lys Ala Glu Leu Gly Ala Leu Pro Asp Asp Phe Ile Asp Ser  
 230 235 240  
 Leu Glu Lys Thr Asp Asp Asp Lys Tyr Lys Ile Thr Leu Lys Tyr  
 245 250 255  
 Pro His Tyr Phe Pro Val Met Lys Lys Cys Cys Ile Pro Glu Thr  
 260 265 270  
 Arg Arg Arg Met Glu Met Ala Phe Asn Thr Arg Cys Lys Glu Glu  
 275 280 285  
 Asn Thr Ile Ile Leu Gln Gln Leu Leu Pro Leu Arg Thr Lys Val  
 290 295 300  
 Ala Lys Leu Leu Gly Tyr Ser Thr His Ala Asp Phe Val Leu Glu  
 305 310 315  
 Met Asn Thr Ala Lys Ser Thr Ser Arg Val Thr Ala Phe Leu Asp  
 320 325 330  
 Asp Leu Ser Gln Lys Leu Lys Pro Leu Gly Glu Ala Glu Arg Glu  
 335 340 345

Phe Ile Leu Asn Leu Lys Lys Glu Cys Lys Asp Arg Gly Phe  
 350 355 360  
 Glu Tyr Asp Gly Lys Ile Asn Ala Trp Asp Leu Tyr Tyr Tyr Met  
 365 370 375  
 Thr Gln Thr Glu Glu Leu Lys Tyr Ser Ile Asp Gln Glu Phe Leu  
 380 385 390  
 Lys Glu Tyr Phe Pro Ile Glu Val Val Thr Glu Gly Leu Leu Asn  
 395 400 405  
 Thr Tyr Gln Glu Leu Leu Gly Leu Ser Phe Glu Gln Met Thr Asp  
 410 415 420  
 Ala His Val Trp Asn Lys Ser Val Thr Leu Tyr Thr Val Lys Asp  
 425 430 435  
 Lys Ala Thr Gly Glu Val Leu Gly Gln Phe Tyr Leu Asp Leu Tyr  
 440 445 450  
 Pro Arg Glu Gly Lys Tyr Asn His Ala Ala Cys Phe Gly Leu Gln  
 455 460 465  
 Pro Gly Cys Leu Leu Pro Asp Gly Ser Arg Met Met Ala Val Ala  
 470 475 480  
 Ala Leu Val Val Asn Phe Ser Gln Pro Val Ala Gly Arg Pro Ser  
 485 490 495  
 Leu Leu Arg His Asp Glu Val Arg Thr Tyr Phe His Glu Phe Gly  
 500 505 510  
 His Val Met His Gln Ile Cys Ala Gln Thr Asp Phe Ala Arg Phe  
 515 520 525  
 Ser Gly Thr Asn Val Glu Thr Asp Phe Val Glu Val Pro Ser Gln  
 530 535 540  
 Met Leu Glu Asn Trp Val Trp Asp Val Asp Ser Leu Arg Arg Leu  
 545 550 555  
 Ser Lys His Tyr Lys Asp Gly Ser Pro Ile Ala Asp Asp Leu Leu  
 560 565 570  
 - Glu-Lys-Leu-Val-Ala-Ser-Arg-Leu-Val-Asn-Thr-Gly-Leu-Leu-Thr  
 575 580 585  
 --Leu-Arg-Gln-Ile-Val-Leu-Ser-Lys-Val-Asp-Gln-Ser-Leu-His-Thr  
 590 595 600  
 Asn-Thr-Ser-Leu-Asp-Ala-Ala-Ser-Glu-Tyr-Ala-Lys-Tyr-Cys-Ser  
 605 610 615  
 -Glu-Ile-Leu-Gly-Val-Ala-Ala-Thr-Pro-Gly-Thr-Asn-Met-Pro-Ala  
 620 625 630  
 -Thr-Phe-Gly-His-Leu-Ala-Gly-Gly-Tyr-Asp-Gly-Gln-Tyr-Tyr-Gly  
 635 640 645  
 -Tyr-Leu-Trp-Ser-Glu-Val-Phe-Ser-Met-Asp-Met-Phe-Tyr-Ser-Cys  
 650 655 660  
 -Phe-Lys-Lys-Glu-Gly-Ile-Met-Asn-Pro-Glu-Val-Gly-Met-Lys-Tyr  
 665 670 675  
 Arg-Asn-Leu-Ile-Leu-Lys-Pro-Gly-Gly-Ser-Leu-Asp-Gly-Met-Asp  
 680 685 690  
 Met-Leu-His-Asn-Phe-Leu-Lys-Arg-Glu-Pro-Asn-Gln-Lys-Ala-Phe  
 695 700 705  
 Leu-Met-Ser-Arg-Gly-Leu-His-Ala-Pro  
 710

<210> 5  
<211> 367  
<212> PRT  
<213> Homo sapiens

<220>  
<221> misc\_feature  
<223> Incyte ID No: 5544089CD1

<400> 5  
Met Phe Ala Pro Ser Val Leu Ser Ser Gly Leu Ser Gly Gly Ala  
 1 5 10 15  
 Ser Lys Gly Arg Lys Met Glu Leu Ile Gln Pro Lys Glu Pro Thr  
 20 25 30  
 Ser Gln Tyr Ile Ser Leu Cys His Glu Leu His Thr Leu Phe Gln  
 35 40 45

Val Met Trp Ser Gly Lys Trp Ala Leu Val Ser Pro Phe Ala Met  
     50                         55                         60  
 Leu His Ser Val Trp Arg Leu Ile Pro Ala Phe Arg Gly Tyr Ala  
     65                         70                         75  
 Gln Gln Asp Ala Gln Glu Phe Leu Cys Glu Leu Leu Asp Lys Ile  
     80                         85                         90  
 Gln Arg Glu Leu Glu Thr Thr Gly Thr Ser Leu Pro Ala Leu Ile  
     95                         100                         105  
 Pro Thr Ser Gln Arg Lys Leu Ile Lys Gln Val Leu Asn Val Val  
   110                         115                         120  
 Asn Asn Ile Phe His Gly Gln Leu Leu Ser Gln Val Thr Cys Leu  
   125                         130                         135  
 Ala Cys Asp Asn Lys Ser Asn Thr Ile Glu Pro Phe Trp Asp Leu  
   140                         145                         150  
 Ser Leu Glu Phe Pro Glu Arg Tyr Gln Cys Ser Gly Lys Asp Ile  
   155                         160                         165  
 Ala Ser Gln Pro Cys Leu Val Thr Glu Met Leu Ala Lys Phe Thr  
   170                         175                         180  
 Glu Thr Glu Ala Leu Glu Gly Lys Ile Tyr Val Cys Asp Gln Cys  
   185                         190                         195  
 Asn Ser Lys Arg Arg Arg Phe Ser Ser Lys Pro Val Val Leu Thr  
   200                         205                         210  
 Glu Ala Gln Lys Gln Leu Met Ile Cys His Leu Pro Gln Val Leu  
   215                         220                         225  
 Arg Leu His Leu Lys Arg Phe Arg Trp Ser Gly Arg Asn Asn Arg  
   230                         235                         240  
 Glu Lys Ile Gly Val His Val Gly Phe Glu Glu Ile Leu Asn Met  
   245                         250                         255  
 Glu Pro Tyr Cys Cys Arg Glu Thr Leu Lys Ser Leu Arg Pro Glu  
   260                         265                         270  
 Cys Phe Ile Tyr Asp Leu Ser Ala Val Val Met His His Gly Lys  
   275                         280                         285  
 Gly Phe Gly Ser Gly His Tyr Thr Ala Tyr Cys Tyr Asn Ser Glu  
   290                         295                         300  
 Gly Gly Phe Trp Val His Cys Asn Asp Ser Lys Leu Ser Met Cys  
   305                         310                         315  
 Thr Met Asp Glu Val Cys Lys Ala Gln Ala Tyr Ile Leu Phe Tyr  
   320                         325                         330  
 Thr Gln Arg Val Thr Glu Asn Gly His Ser Lys Leu Leu Pro Pro  
   335                         340                         345  
 Glu Leu Leu Leu Gly Ser Gln His Pro Asn Glu Asp Ala Asp Thr  
   350                         355                         360  
 Ser Ser Asn Glu Ile Leu Ser  
   365

<210> 6  
 <211> 235  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 7474081CD1

<400> 6

Met	Lys	Tyr	Val	Phe	Tyr	Leu	Gly	Val	Leu	Ala	Gly	Thr	Phe	Phe
1			5				10				15			
Phe	Ala	Asp	Ser	Ser	Val	Gln	Lys	Glu	Asp	Pro	Ala	Pro	Tyr	Leu
	20						25						30	
Val	Tyr	Leu	Lys	Ser	His	Phe	Asn	Pro	Cys	Val	Gly	Val	Leu	Ile
	35						40						45	
Lys	Pro	Ser	Trp	Val	Leu	Ala	Pro	Ala	His	Cys	Tyr	Leu	Pro	Asn
	50						55						60	
Leu	Lys	Val	Met	Leu	Gly	Asn	Phe	Lys	Ser	Arg	Val	Arg	Asp	Gly
	65						70						75	
Thr	Glu	Gln	Thr	Ile	Asn	Pro	Ile	Gln	Ile	Val	Arg	Tyr	Trp	Asn
	80						85						90	

Tyr Ser His Ser Ala Pro Gln Asp Asp Leu Met Leu Ile Lys Leu  
                   95                   100                   105  
 Ala Lys Pro Ala Met Leu Asn Pro Lys Val Gln Pro Leu Thr Leu  
                   110                   115                   120  
 Ala Thr Thr Asn Val Arg Pro Gly Thr Val Cys Leu Leu Ser Gly  
                   125                   130                   135  
 Leu Asp Trp Ser Gln Glu Asn Ser Gly Arg His Pro Asp Leu Arg  
                   140                   145                   150  
 Gln Asn Leu Glu Ala Pro Val Met Ser Asp Arg Glu Cys Gln Lys  
                   155                   160                   165  
 Thr Glu Gln Gly Lys Ser His Arg Asn Ser Leu Cys Val Lys Phe  
                   170                   175                   180  
 Val Lys Val Phe Ser Arg Ile Phe Gly Glu Val Ala Val Ala Thr  
                   185                   190                   195  
 Val Ile Cys Lys Asp Lys Leu Gln Gly Ile Glu Val Gly His Phe  
                   200                   205                   210  
 Met Gly Gly Asp Val Gly Ile Tyr Thr Asn Val Tyr Lys Tyr Val  
                   215                   220                   225  
 Ser Trp Ile Glu Asn Thr Ala Lys Asp Lys  
                   230                   235

&lt;210&gt; 7

&lt;211&gt; 488

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 5281209CD1

&lt;400&gt; 7

Met Gln Pro Thr Gly Arg Glu Gly Ser Arg Ala Leu Ser Arg Arg  
     1               5               10               15  
 Tyr Leu Arg Arg Leu Leu Leu Leu Leu Leu Leu Leu Leu Arg  
     20               25               30  
 Gln Pro Val Thr Arg Ala Glu Thr Thr Pro Gly Ala Pro Arg Ala  
     35               40               45  
 Leu Ser Thr Leu Gly Ser Pro Ser Leu Phe Thr Thr Pro Gly Val  
     50               55               60  
 Pro Ser Ala Leu Thr Pro Gly Leu Thr Thr Pro Gly Thr Pro  
     65               70               75  
 Lys Thr Leu Asp Leu Arg Gly Arg Ala Gln Ala Leu Met Arg Ser  
     80               85               90  
 Phe Pro Leu Val Asp Gly His Asn Asp Leu Pro Gln Val Leu Arg  
     95               100               105  
 Gln Arg Tyr Lys Asn Val Leu Gln Asp Val Asn Leu Arg Asn Phe  
   110               115               120  
 Ser His Gly Gln Thr Ser Leu Asp Arg Leu Arg Asp Gly Leu Val  
   125               130               135  
 Gly Ala Gln Phe Trp Ser Ala Ser Val Ser Cys Gln Ser Gln Asp  
   140               145               150  
 Gln Thr Ala Val Arg Leu Ala Leu Glu Gln Ile Asp Leu Ile His  
   155               160               165  
 Arg Met Cys Ala Ser Tyr Ser Glu Leu Glu Leu Val Thr Ser Ala  
   170               175               180  
 Glu Gly Leu Asn Ser Ser Gln Lys Leu Ala Cys Leu Ile Gly Val  
   185               190               195  
 Glu Gly Gly His Ser Leu Asp Ser Ser Leu Ser Val Leu Arg Ser  
   200               205               210  
 Phe Tyr Val Leu Gly Val Arg Tyr Leu Thr Leu Thr Phe Thr Cys  
   215               220               225  
 Ser Thr Pro Trp Ala Glu Ser Ser Thr Lys Phe Arg His His Met  
   230               235               240  
 Tyr Thr Asn Val Ser Gly Leu Thr Ser Phe Gly Glu Lys Val Val  
   245               250               255  
 Glu Glu Leu Asn Arg Leu Gly Met Met Ile Asp Leu Ser Tyr Ala  
   260               265               270

<210> 8  
<211> 346  
<212> PRT  
<213> *Homo sapiens*

<220>  
<221> misc\_feature  
<223> Incyte ID No: 2256251cd1

<400>	8													
Met	Leu	Leu	Gly	Arg	Val	Trp	Gln	Thr	Arg	Glu	Leu	Lys	Ser	Lys
1					5				10					15
Val	Pro	Lys	Lys	Ala	Gly	Arg	Cys	Gly	Gln	Gly	Arg	Leu	His	Gly
					20				25					30
Gly	Ser	Ala	Val	Gly	Phe	Leu	Gly	Ser	Pro	Pro	Gly	Thr	Pro	Ser
					35				40					45
Ser	Phe	Asp	Leu	Gly	Cys	Gly	Arg	Pro	Gln	Val	Ser	Asp	Ala	Gly
					50				55					60
Gly	Arg	Ile	Val	Gly	Gly	His	Ala	Ala	Pro	Ala	Gly	Ala	Trp	Pro
					65				70					75
Trp	Gln	Ala	Ser	Leu	Arg	Leu	Arg	Arg	Val	His	Val	Cys	Gly	Gly
					80				85					90
Ser	Leu	Leu	Ser	Pro	Gln	Trp	Val	Leu	Thr	Ala	Ala	His	Cys	Phe
				95					100					105
Ser	Gly	Ser	Leu	Asn	Ser	Ser	Asp	Tyr	Gln	Val	His	Leu	Gly	Glut
				110					115					120
Leu	Glu	Ile	Thr	Leu	Ser	Pro	His	Phe	Ser	Thr	Val	Arg	Gln	Ile
				125					130					135
Ile	Leu	His	Ser	Ser	Pro	Ser	Gly	Gln	Pro	Gly	Thr	Ser	Gly	Asp
				140					145					150
Ile	Ala	Leu	Val	Glu	Leu	Ser	Val	Pro	Val	Thr	Leu	Phe	Ser	Arg
				155					160					165
Ile	Leu	Pro	Val	Cys	Leu	Pro	Gl	Ala	Ser	Asp	Asp	Phe	Cys	Pro
				170					175					180
Gly	Ile	Arg	Cys	Trp	Val	Thr	Gly	Trp	Gly	Tyr	Thr	Arg	Glu	Gly
				185					190					195

Glu	Pro	Leu	Pro	Pro	Pro	Tyr	Ser	Leu	Arg	Glu	Val	Lys	Val	Ser
200								205					210	
Val	Val	Asp	Thr	Glu	Thr	Cys	Arg	Arg	Asp	Tyr	Pro	Gly	Pro	Gly
215								220					225	
Gly	Ser	Ile	Leu	Gln	Pro	Asp	Met	Leu	Cys	Ala	Arg	Gly	Pro	Gly
230								235					240	
Asp	Ala	Cys	Gln	Asp	Asp	Ser	Gly	Gly	Pro	Leu	Val	Cys	Gln	Val
245								250					255	
Asn	Gly	Ala	Trp	Val	Gln	Ala	Gly	Ile	Val	Ser	Trp	Gly	Glu	Gly
260								265					270	
Cys	Gly	Arg	Pro	Asn	Arg	Pro	Gly	Val	Tyr	Thr	Arg	Val	Pro	Ala
275								280					285	
Tyr	Val	Asn	Trp	Ile	Arg	Arg	His	Ile	Thr	Ala	Ser	Gly	Gly	Ser
290								295					300	
Glu	Ser	Gly	Tyr	Pro	Arg	Leu	Pro	Leu	Leu	Ala	Gly	Leu	Phe	Leu
305								310					315	
Pro	Gly	Leu	Phe	Leu	Leu	Leu	Val	Ser	Cys	Val	Leu	Leu	Ala	Lys
320								325					330	
Cys	Leu	Leu	His	Pro	Ser	Ala	Asp	Gly	Thr	Pro	Phe	Pro	Ala	Pro
335								340					345	
Asp														

<210> 9  
<211> 882  
<212> PRT  
<213> Homo sapiens

<220>  
<221> misc\_feature  
<223> Incyte ID No: 7160544CD1

<400> 9  
Met Ala Ala Ala Met Glu Thr Glu Gln Leu Gly Val Glu Ile Phe  
1 5 10 15  
Glu Thr Ala Asp Cys Glu Glu Asn Ile Glu Ser Gln Asp Arg Pro  
20 25 30  
Lys Leu Glu Pro Phe Tyr Val Glu Arg Tyr Ser Trp Ser Gln Leu  
35 40 45  
Lys Lys Leu Ala Asp Thr Arg Lys Tyr His Gly Tyr Met Met  
50 55 60  
Ala Lys Ala Pro His Asp Phe Met Phe Val Lys Arg Asn Asp Pro  
65 70 75  
Asp Gly Pro His Ser Asp Arg Ile Tyr Tyr Leu Ala Met Ser Gly  
80 85 90  
Glu Asn Arg Glu Asn Thr Leu Phe Tyr Ser Glu Ile Pro Lys Thr  
95 100 105  
Ile Asn Arg Ala Ala Val Leu Met Leu Ser Trp Lys Pro Leu Leu  
110 115 120  
Asp Leu Phe Gln Ala Thr Leu Asp Tyr Gly Met Tyr Ser Arg Glu  
125 130 135  
Glu Glu Leu Leu Arg Glu Arg Lys Arg Ile Gly Thr Val Gly Ile  
140 145 150  
Ala Ser Tyr Asp Tyr His Gln Gly Ser Gly Thr Phe Leu Phe Gln  
155 160 165  
Ala Gly Ser Gly Ile Tyr His Val Lys Asp Gly Gly Pro Gln Gly  
170 175 180  
Phe Thr Gln Gln Pro Leu Arg Pro Asn Leu Val Glu Thr Ser Cys  
185 190 195  
Pro Asn Ile Arg Met Asp Pro Lys Leu Cys Pro Ala Asp Pro Asp  
200 205 210  
Trp Ile Ala Phe Ile His Ser Asn Asp Ile Trp Ile Ser Asn Ile  
215 220 225  
Val Thr Arg Glu Glu Arg Arg Leu Thr Tyr Val His Asn Glu Leu  
230 235 240  
Ala Asn Met Glu Glu Asp Ala Arg Ser Ala Gly Val Ala Thr Phe  
245 250 255

Val Leu Gln Glu Glu Phe Asp Arg Tyr Ser Gly Tyr Trp Trp Cys  
 260 265 270  
 Pro Lys Ala Glu Thr Thr Pro Ser Gly Gly Lys Ile Leu Arg Ile  
 275 280 285  
 Leu Tyr Glu Glu Asn Asp Glu Ser Glu Val Glu Ile Ile His Val  
 290 295 300  
 Thr Ser Pro Met Leu Glu Thr Arg Arg Ala Asp Ser Phe Arg Tyr  
 305 310 315  
 Pro Lys Thr Gly Thr Ala Asn Pro Lys Val Thr Phe Lys Met Ser  
 320 325 330  
 Glu Ile Met Ile Asp Ala Glu Gly Arg Ile Ile Asp Val Ile Asp  
 335 340 345  
 Lys Glu Leu Ile Gln Pro Phe Glu Ile Leu Phe Glu Gly Val Glu  
 350 355 360  
 Tyr Ile Ala Arg Ala Gly Trp Thr Pro Glu Gly Lys Tyr Ala Trp  
 365 370 375  
 Ser Ile Leu Leu Asp Arg Ser Gln Thr Arg Leu Gln Ile Val Leu  
 380 385 390  
 Ile Ser Pro Glu Leu Phe Ile Pro Val Glu Asp Asp Val Met Glu  
 395 400 405  
 Arg Gln Arg Leu Ile Glu Ser Val Pro Asp Ser Val Thr Pro Leu  
 410 415 420  
 Ile Ile Tyr Glu Glu Thr Thr Asp Ile Trp Ile Asn Ile His Asp  
 425 430 435  
 Ile Phe His Val Phe Pro Gln Ser His Glu Glu Glu Ile Glu Phe  
 440 445 450  
 Ile Phe Ala Ser Glu Cys Lys Thr Gly Phe Arg His Leu Tyr Lys  
 455 460 465  
 Ile Thr Ser Ile Leu Lys Glu Ser Lys Tyr Lys Arg Ser Ser Gly  
 470 475 480  
 Gly Leu Pro Ala Pro Ser Asp Phe Lys Cys Pro Ile Lys Glu Glu  
 485 490 495  
 Ile Ala Ile Thr Ser Gly Glu Trp Glu Val Leu Gly Arg His Gly  
 500 505 510  
 Ser Asn Ile Gln Val Asp Glu Val Arg Arg Leu Val Tyr Phe Glu  
 515 520 525  
 Gly Thr Lys Asp Ser Pro Leu Glu His His Leu Tyr Val Val Ser  
 530 535 540  
 Tyr Val Asn Pro Gly Glu Val Thr Arg Leu Thr Asp Arg Gly Tyr  
 545 550 555  
 Ser His Ser Cys Cys Ile Ser Gln His Cys Asp Phe Phe Ile Ser  
 560 565 570  
 Lys Tyr Ser Asn Gln Lys Asn Pro His Cys Val Ser Leu Tyr Lys  
 575 580 585  
 Leu Ser Ser Pro Glu Asp Asp Pro Thr Cys Lys Thr Lys Glu Phe  
 590 595 600  
 Trp Ala Thr Ile Leu Asp Ser Ala Gly Pro Leu Pro Asp Tyr Thr  
 605 610 615  
 Pro Pro Glu Ile Phe Ser Phe Glu Ser Thr Thr Gly Phe Thr Leu  
 620 625 630  
 Tyr Gly Met Leu Tyr Lys Pro His Asp Leu Gln Pro Gly Lys Lys  
 635 640 645  
 Tyr Pro Thr Val Leu Phe Ile Tyr Gly Gly Pro Gln Val Gln Leu  
 650 655 660  
 Val Asn Asn Arg Phe Lys Gly Val Lys Tyr Phe Arg Leu Asn Thr  
 665 670 675  
 Leu Ala Ser Leu Gly Tyr Val Val Val Val Ile Asp Asn Arg Gly  
 680 685 690  
 Ser Cys His Arg Gly Leu Lys Phe Glu Gly Ala Phe Lys Tyr Lys  
 695 700 705  
 Met Gly Gln Ile Glu Ile Asp Asp Gln Val Glu Gly Leu Gln Tyr  
 710 715 720  
 Leu Ala Ser Arg Tyr Asp Phe Ile Asp Leu Asp Arg Val Gly Ile  
 725 730 735  
 His Gly Trp Ser Tyr Gly Gly Tyr Leu Ser Leu Met Ala Leu Met  
 740 745 750  
 Gln Arg Ser Asp Ile Phe Arg Val Ala Ile Ala Gly Ala Pro Val

755	760	765
Thr Leu Trp Ile Phe Tyr Asp Thr Gly	Tyr Thr Glu Arg Tyr	Met
770	775	780
Gly His Pro Asp Gln Asn Glu Gln Gly	Tyr Tyr Leu Gly Ser	Val
785	790	795
Ala Met Gln Ala Glu Lys Phe Pro Ser	Glu Pro Asn Arg Leu	Leu
800	805	810
Leu Leu His Gly Phe Leu Asp Glu Asn	Val His Phe Ala His	Thr
815	820	825
Ser Ile Leu Leu Ser Phe Leu Val Arg	Ala Gly Lys Pro Tyr	Asp
830	835	840
Leu Gln Ile Tyr Pro Gln Glu Arg His	Ser Ile Arg Val Pro	Glu
845	850	855
Ser Gly Glu His Tyr Glu Leu His Leu	Leu His Tyr Leu Gln	Glu
860	865	870
Asn Leu Gly Ser Arg Ile Ala Ala Leu	Lys Val Ile	
875	880	

<210> 10  
<211> 1189  
<212> PRT  
<213> Homo sapiens

<220>  
<221> misc\_feature  
<223> Incyte ID No: 7477386CD1

<400> 10  
Met Ala Pro Leu Arg Ala Leu Leu Ser Tyr Leu Leu Pro Leu His  
1 5 10 15  
Cys Ala Leu Cys Ala Ala Ala Gly Ser Arg Thr Pro Glu Leu His  
20 25 30  
Leu Ser Gly Lys Leu Ser Asp Tyr Gly Val Thr Val Pro Cys Ser  
35 40 45  
Thr Asp Phe Arg Gly Arg Phe Leu Ser His Val Val Ser Gly Pro  
50 55 60  
Ala Ala Ala Ser Ala Gly Ser Met Val Val Asp Thr Pro Pro Thr  
65 70 75  
Leu Pro Arg His Ser Ser His Leu Arg Val Ala Arg Ser Pro Leu  
80 85 90  
His Pro Gly Gly Thr Leu Trp Pro Gly Arg Val Gly Arg His Ser  
95 100 105  
Leu Tyr Phe Asn Val Thr Val Phe Gly Lys Glu Leu His Leu Arg  
110 115 120  
Leu Arg Pro Asn Arg Arg Leu Val Val Pro Gly Ser Ser Val Glu  
125 130 135  
Trp Gln Glu Asp Phe Arg Glu Leu Phe Arg Gln Pro Leu Arg Gln  
140 145 150  
Glu Cys Val Tyr Thr Gly Gly Val Thr Gly Met Pro Gly Ala Ala  
155 160 165  
Val Ala Ile Ser Asn Cys Asp Gly Leu Ala Gly Leu Ile Arg Thr  
170 175 180  
Asp Ser Thr Asp Phe Phe Ile Glu Pro Leu Glu Arg Gly Gln Gln  
185 190 195  
Glu Lys Glu Ala Ser Gly Arg Thr His Val Val Tyr Arg Arg Glu  
200 205 210  
Ala Val Gln Gln Glu Trp Ala Glu Pro Asp Gly Asp Leu His Asn  
215 220 225  
Glu Ala Phe Gly Leu Gly Asp Leu Pro Asn Leu Leu Gly Leu Val  
230 235 240  
Gly Asp Gln Leu Gly Asp Thr Glu Arg Lys Arg Arg His Ala Lys  
245 250 255  
Pro Gly Ser Tyr Ser Ile Glu Val Leu Leu Val Val Asp Asp Ser  
260 265 270  
Val Val Arg Phe His Gly Lys Glu His Val Gln Asn Tyr Val Leu  
275 280 285  
Thr Leu Met Asn Ile Val Val Asp Glu Ile Tyr His Asp Glu Ser

	290	295	300											
Leu	Gly	Val	His	Ile	Asn	Ile	Ala	Leu	Val	Arg	Leu	Ile	Met	Val
				305					310					315
Gly	Tyr	Arg	Gln	Gln	Ser	Leu	Ser	Leu	Ile	Glu	Arg	Gly	Asn	Pro
									320					330
Ser	Arg	Ser	Leu	Glu	Gln	Val	Cys	Arg	Trp	Ala	His	Ser	Gln	Gln
									335					345
Arg	Gln	Asp	Pro	Ser	His	Ala	Glu	His	His	Asp	His	Val	Val	Phe
									350					360
Leu	Thr	Arg	Gln	Asp	Phe	Gly	Pro	Ser	Gly	Gly	Tyr	Ala	Pro	Val
									365					375
Thr	Gly	Met	Cys	His	Pro	Leu	Arg	Ser	Cys	Ala	Leu	Asn	His	Glu
									380					390
Asp	Gly	Phe	Ser	Ser	Ala	Phe	Val	Ile	Ala	His	Glu	Thr	Gly	His
									395					405
Val	Leu	Gly	Met	Glu	His	Asp	Gly	Gln	Gly	Asn	Gly	Cys	Ala	Asp
									410					420
Glu	Thr	Ser	Leu	Gly	Ser	Val	Met	Ala	Pro	Leu	Val	Gln	Ala	Ala
									425					435
Phe	His	Arg	Phe	His	Trp	Ser	Arg	Cys	Ser	Lys	Leu	Glu	Leu	Ser
									440					450
Arg	Tyr	Leu	Pro	Ser	Tyr	Asp	Cys	Leu	Leu	Asp	Asp	Pro	Phe	Asp
									455					465
Pro	Ala	Trp	Pro	Gln	Pro	Pro	Glu	Leu	Pro	Gly	Ile	Asn	Tyr	Ser
									470					480
Met	Asp	Glu	Gln	Cys	Arg	Phe	Asp	Phe	Gly	Ser	Gly	Tyr	Gln	Thr
									485					495
Cys	Leu	Ala	Phe	Arg	Thr	Phe	Glu	Pro	Cys	Lys	Gln	Leu	Trp	Cys
									500					510
Ser	His	Pro	Asp	Asn	Pro	Tyr	Phe	Cys	Lys	Thr	Lys	Lys	Gly	Pro
									515					525
Pro	Leu	Asp	Gly	Thr	Glu	Cys	Ala	Pro	Gly	Lys	Trp	Cys	Phe	Lys
									530					540
Gly	His	Cys	Ile	Trp	Lys	Ser	Pro	Glu	Gln	Thr	Tyr	Gly	Gln	Asp
									545					555
Gly	Gly	Trp	Ser	Ser	Trp	Thr	Lys	Phe	Gly	Ser	Cys	Ser	Arg	Ser
									560					570
Cys	Gly	Gly	Gly	Val	Arg	Ser	Arg	Ser	Arg	Ser	Cys	Asn	Asn	Pro
									575					585
Ser	Pro	Ala	Tyr	Gly	Gly	Arg	Leu	Cys	Leu	Gly	Pro	Met	Phe	Glu
									590					600
Tyr	Gln	Val	Cys	Asn	Ser	Glu	Glu	Cys	Pro	Gly	Thr	Tyr	Glu	Asp
									605					615
Phe	Arg	Ala	Gln	Gln	Cys	Ala	Lys	Arg	Asn	Ser	Tyr	Tyr	Val	His
									620					630
Gln	Asn	Ala	Lys	His	Ser	Trp	Val	Pro	Tyr	Glu	Pro	Asp	Asp	Asp
									635					645
Ala	Gln	Lys	Cys	Glu	Leu	Ile	Cys	Gln	Ser	Ala	Asp	Thr	Gly	Asp
									650					660
Val	Val	Phe	Met	Asn	Gln	Val	Val	His	Asp	Gly	Thr	Arg	Cys	Ser
									665					675
Tyr	Arg	Asp	Pro	Tyr	Ser	Val	Cys	Ala	Arg	Gly	Glu	Cys	Val	Pro
									680					690
Val	Gly	Cys	Asp	Lys	Glu	Val	Gly	Ser	Met	Lys	Ala	Asp	Asp	Lys
									695					705
Cys	Gly	Val	Cys	Gly	Gly	Asp	Asn	Ser	His	Cys	Arg	Thr	Val	Lys
									710					720
Gly	Thr	Leu	Gly	Lys	Ala	Ser	Lys	Gln	Ala	Gly	Ala	Leu	Lys	Leu
									725					735
Val	Gln	Ile	Pro	Ala	Gly	Ala	Arg	His	Ile	Gln	Ile	Glu	Ala	Leu
									740					750
Glu	Lys	Ser	Pro	His	Arg	Ile	Val	Val	Lys	Asn	Gln	Val	Thr	Gly
									755					765
Ser	Phe	Ile	Leu	Asn	Pro	Lys	Gly	Lys	Glu	Ala	Thr	Ser	Arg	Thr
									770					780
Phe	Thr	Ala	Met	Gly	Leu	Glu	Trp	Glu	Asp	Ala	Val	Glu	Asp	Ala
									785					795

Lys Glu Ser Leu Lys Thr Ser Gly Pro Leu Pro Glu Ala Ile Ala  
 800 805 810  
 Ile Leu Ala Leu Pro Pro Thr Glu Gly Gly Pro Arg Ser Ser Leu  
 815 820 825  
 Ala Tyr Lys Tyr Val Ile His Glu Asp Leu Leu Pro Leu Ile Gly  
 830 835 840  
 Ser Asn Asn Val Leu Leu Glu Glu Met Asp Thr Tyr Glu Trp Ala  
 845 850 855  
 Leu Lys Ser Trp Ala Pro Cys Ser Lys Ala Cys Gly Gly Gly Ile  
 860 865 870  
 Gln Phe Thr Lys Tyr Gly Cys Arg Arg Arg Arg Asp His His Met  
 875 880 885  
 Val Gln Arg His Leu Cys Asp His Lys Lys Arg Pro Lys Pro Ile  
 890 895 900  
 Arg Arg Arg Cys Asn Gln His Pro Cys Ser Gln Pro Val Trp Val  
 905 910 915  
 Thr Glu Glu Trp Gly Ala Cys Ser Arg Ser Cys Gly Lys Leu Gly  
 920 925 930  
 Val Gln Thr Arg Gly Ile Gln Cys Leu Leu Pro Leu Ser Asn Gly  
 935 940 945  
 Thr His Lys Val Met Pro Ala Lys Ala Cys Ala Gly Asp Arg Pro  
 950 955 960  
 Glu Ala Arg Arg Pro Cys Leu Arg Val Pro Cys Pro Ala Gln Trp  
 965 970 975  
 Arg Leu Gly Ala Trp Ser Gln Cys Ser Ala Thr Cys Gly Glu Gly  
 980 985 990  
 Ile Gln Gln Arg Gln Val Val Cys Arg Thr Asn Ala Asn Ser Leu  
 995 1000 1005  
 Gly-His-Cys-Glu-Gly Asp Arg Pro Asp Thr Val Gln Val Cys Ser  
 1010 1015 1020  
 Leu-Pro-Ala-Cys-Gly-Ala-Glu-Pro-Cys-Thr-Gly-Asp-Arg-Ser-Val  
 1025 1030 1035  
 Phe-Gly-Gln-Met-Glu-Val-Leu-Asp-Arg-Tyr-Cys-Ser-Ile-Pro-Gly  
 1040 1045 1050  
 Tyr-His-Arg-Leu-Cys-Cys-Val-Ser-Cys-Ile-Lys-Lys-Ala-Ser-Gly  
 1055 1060 1065  
 Pro-Asn-Pro-Gly-Pro-Asp-Pro-Gly-Pro-Thr-Ser-Leu-Pro-Pro-Phe  
 1070 1075 1080  
 Ser-Thr-Pro-Gly-Ser-Pro-Leu-Pro-Gly-Pro-Gln-Asp-Pro-Ala-Asp  
 1085 1090 1095  
 Ala-Ala-Glu-Pro-Pro-Gly-Lys-Pro-Thr-Gly-Ser-Glu-Asp-His-Gln  
 1100 1105 1110  
 His-Gly-Arg-Ala-Thr-Gln-Leu-Pro-Gly-Ala-Leu-Asp-Thr-Ser-Ser  
 1115 1120 1125  
 Pro-Gly-Thr-Gln-His-Pro-Phe-Ala-Pro-Glu-Thr-Pro-Ile-Pro-Gly  
 1130 1135 1140  
 Ala-Ser-Trp-Ser-Ile-Ser-Pro-Thr-Thr-Pro-Gly-Gly-Leu-Pro-Trp  
 1145 1150 1155  
 Gly-Trp-Thr-Gln-Thr-Pro-Thr-Pro-Val-Pro-Glu-Asp-Lys-Gly-Gln  
 1160 1165 1170  
 Pro-Gly-Glu-Asp-Leu-Arg-His-Pro-Gly-Thr-Ser-Leu-Pro-Ala-Ala  
 1175 1180 1185  
 Ser-Pro-Val-Thr

<210> 11  
 <211> 952  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 7473089CD1

<400> 11  
 Met Leu Leu Leu Gly Ile Leu Thr Leu Ala Phe Ala Gly Arg Thr  
 1 5 10 15

Ala Gly Gly Ser Glu Pro Glu Arg Glu Val Val Val Pro Ile Arg  
   20                   25                   30  
 Leu Asp Pro Asp Ile Asn Gly Arg Arg Tyr Tyr Trp Arg Gly Pro  
   35                   40                   45  
 Glu Asp Ser Gly Asp Gln Gly Leu Ile Phe Gln Ile Thr Ala Phe  
   50                   55                   60  
 Gln Glu Asp Phe Tyr Leu His Leu Thr Pro Asp Ala Gln Phe Leu  
   65                   70                   75  
 Ala Pro Ala Phe Ser Thr Glu His Leu Gly Val Pro Leu Gln Gly  
   80                   85                   90  
 Leu Thr Gly Gly Ser Ser Asp Leu Arg Arg Cys Phe Tyr Ser Gly  
   95                   100                  105  
 Asp Val Asn Ala Glu Pro Asp Ser Phe Ala Ala Val Ser Leu Cys  
 110                   115                  120  
 Gly Gly Leu Arg Gly Ala Phe Gly Tyr Arg Gly Ala Glu Tyr Val  
 125                   130                  135  
 Ile Ser Pro Leu Pro Asn Ala Ser Ala Pro Ala Ala Gln Arg Asn  
 140                   145                  150  
 Ser Gln Gly Ala His Leu Leu Gln Arg Arg Gly Val Pro Gly Gly  
 155                   160                  165  
 Pro Ser Gly Asp Pro Thr Ser Arg Cys Gly Val Ala Ser Gly Trp  
 170                   175                  180  
 Asn Pro Ala Ile Leu Arg Ala Leu Asp Pro Tyr Lys Pro Arg Arg  
 185                   190                  195  
 Ala Gly Phe Gly Glu Ser Arg Ser Arg Arg Arg Ser Gly Arg Ala  
 200                   205                  210  
 Lys Arg Phe Val Ser Ile Pro Arg Tyr Val Glu Thr Leu Val Val  
 215                   220                  225  
 Ala Asp Glu Ser Met Val Lys Phe His Gly Ala Asp Leu Glu His  
 230                   235                  240  
 Tyr Leu Leu Thr Leu Leu Ala Thr Ala Ala Arg Leu Tyr Arg His  
 245                   250                  255  
 Pro Ser Ile Leu Asn Pro Ile Asn Ile Val Val Val Lys Val Leu  
 260                   265                  270  
 Leu Leu Arg Asp Arg Asp Ser Gly Pro Lys Val Thr Gly Asn Ala  
 275                   280                  285  
 Ala Leu Thr Leu Arg Asn Phe Cys Ala Trp Gln Lys Lys Leu Asn  
 290                   295                  300  
 Lys Val Ser Asp Lys His Pro Glu Tyr Trp Asp Thr Ala Ile Leu  
 305                   310                  315  
 Phe Thr Arg Gln Asp Leu Cys Gly Ala Thr Thr Cys Asp Thr Leu  
 320                   325                  330  
 Gly Met Ala Asp Val Gly Thr Met Cys Asp Pro Lys Arg Ser Cys  
 335                   340                  345  
 Ser Val Ile Glu Asp Asp Gly Leu Pro Ser Ala Phe Thr Thr Ala  
 350                   355                  360  
 His Glu Leu Gly His Val Phe Asn Met Pro His Asp Asn Val Lys  
 365                   370                  375  
 Val Cys Glu Glu Val Phe Gly Lys Leu Arg Ala Asn His Met Met  
 380                   385                  390  
 Ser Pro Thr Leu Ile Gln Ile Asp Arg Ala Asn Pro Trp Ser Ala  
 395                   400                  405  
 Cys Ser Ala Ala Ile Ile Thr Asp Phe Leu Asp Ser Gly His Gly  
 410                   415                  420  
 Asp Cys Leu Leu Asp Gln Pro Ser Lys Pro Ile Ser Leu Pro Glu  
 425                   430                  435  
 Asp Leu Pro Gly Ala Ser Tyr Thr Leu Ser Gln Gln Cys Glu Leu  
 440                   445                  450  
 Ala Phe Gly Val Gly Ser Lys Pro Cys Pro Tyr Met Gln Tyr Cys  
 455                   460                  465  
 Thr Lys Leu Trp Cys Thr Gly Lys Ala Lys Gly Gln Met Val Cys  
 470                   475                  480  
 Gln Thr Arg His Phe Pro Trp Ala Asp Gly Thr Ser Cys Gly Glu  
 485                   490                  495  
 Gly Lys Leu Cys Leu Lys Gly Ala Cys Val Glu Arg His Asn Leu  
 500                   505                  510  
 Asn Lys His Arg Val Asp Gly Ser Trp Ala Lys Trp Asp Pro Tyr

515		520		525
Gly Pro Cys Ser Arg	Thr Cys Gly Gly	Gly Val Gln Leu Ala	Arg	
530		535		540
Arg Gln Cys Thr Asn	Pro Thr Pro Ala Asn	Gly Gly Lys Tyr	Cys	
545		550		555
Glu Gly Val Arg Val	Lys Tyr Arg Ser	Cys Asn Leu Glu Pro	Cys	
560		565		570
Pro Ser Ser Ala Ser	Gly Lys Ser Phe Arg	Glu Glu Gln Cys	Glu	
575		580		585
Ala Phe Asn Gly Tyr	Asn His Ser Thr Asn	Arg Leu Thr Leu	Ala	
590		595		600
Val Ala Trp Val Pro	Lys Tyr Ser Gly Val	Ser Pro Arg Asp	Lys	
605		610		615
Cys Lys Leu Ile Cys	Arg Ala Asn Gly	Thr Gly Tyr Phe Tyr	Val	
620		625		630
Leu Ala Pro Lys Val	Val Val Asp Gly	Thr Leu Cys Ser Pro	Asp	
635		640		645
Ser Thr Ser Val Cys	Val Gln Gly Lys	Cys Ile Lys Ala Gly	Cys	
650		655		660
Asp Gly Asn Leu Gly	Ser Lys Lys Arg	Phe Asp Lys Cys Gly	Val	
665		670		675
Cys Gly Gly Asp Asn	Lys Ser Cys Lys	Lys Val Thr Gly Leu	Phe	
680		685		690
Thr Lys Pro Met His	Gly Tyr Asn Phe	Val Val Ala Ile Pro	Ala	
695		700		705
Gly Ala Ser Ser Ile	Asp Ile Arg Gln	Arg Gly Tyr Lys Gly	Leu	
710		715		720
Ile Gly Asp Asp Asn	Tyr Leu Ala Leu	Lys Asn Ser Gln Gly	Lys	
725		730		735
*Tyr-Leu-Leu-Asn-Gly	His Phe Val Val	Ser Ala Val Glu Arg	Asp	
740		745		750
*Leu-Val-Val-Lys-Gly-Ser	Leu Leu Arg	Tyr Ser Gly Thr Gly	Thr	
755		760		765
*Ala-Val-Glu-Ser-Leu-Gln-Ala-Ser	Arg Pro Ile Leu Glu Pro	Leu		
770		775		780
*Thr-Val-Glu-Val-Leu-Ser-Val-Gly	Lys Met	Thr Pro Pro Arg	Val	
785		790		795
*Arg-Tyr-Ser-Phe-Tyr-Leu-Pro	Lys Glu Pro	Arg Glu Asp Lys	Ser	
800		805		810
*Ser-His-Pro-Pro-His	Pro Arg Gly Gly	Pro Ser Val Leu His	Asn	
815		820		825
*Ser-Val-Leu-Ser-Leu-Ser-Asn	Gln Val Glu Gln	Pro Asp Asp	Arg	
830		835		840
Pro Pro Ala Arg Trp	Val Ala Gly Ser	Trp Gly Pro Cys Ser	Ala	
845		850		855
Ser Cys Gly Ser Gly	Leu Gln Lys Arg	Ala Val Asp Trp Arg	Gly	
860		865		870
Ser Ala Gly Gln Arg	Thr Val Pro Ala Cys	Asp Ala Ala His	Arg	
875		880		885
Pro Val Glu Thr Gln	Ala Cys Gly Glu	Pro Cys Pro Thr Trp	Glu	
890		895		900
Leu Ser Ala Trp Ser	Pro Cys Ser Lys	Ser Cys Gly Arg Gly	Phe	
905		910		915
Gln Arg Arg Ser Leu	Lys Cys Val Gly	His Gly Gly Arg Leu	Leu	
920		925		930
Ala Arg Asp Gln Cys	Asn Leu His Arg	Lys Pro Gln Glu Leu	Asp	
935		940		945
Phe Cys Val Leu Arg	Pro Cys			
950				

&lt;210&gt; 12

&lt;211&gt; 898

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7604035CD1

&lt;400&gt; 12

Met Glu Asn Trp Thr Gly Arg Pro Trp Leu Tyr Leu Leu Leu Leu  
 1 5 10 15  
 Leu Ser Leu Pro Gln Leu Cys Leu Asp Gln Glu Val Leu Ser Gly  
 20 25 30  
 His Ser Leu Gln Thr Pro Thr Glu Glu Gly Gln Gly Pro Glu Gly  
 35 40 45  
 Val Trp Gly Pro Trp Val Gln Trp Ala Ser Cys Ser Gln Pro Cys  
 50 55 60  
 Gly Val Gly Val Gln Arg Arg Ser Arg Thr Cys Gln Leu Pro Thr  
 65 70 75  
 Val Gln Leu His Pro Ser Leu Pro Leu Pro Pro Arg Pro Pro Arg  
 80 85 90  
 His Pro Glu Ala Leu Leu Pro Arg Gly Gln Gly Pro Arg Pro Gln  
 95 100 105  
 Thr Ser Pro Glu Thr Leu Pro Leu Tyr Arg Thr Gln Ser Arg Gly  
 110 115 120  
 Arg Gly Gly Pro Leu Arg Gly Pro Ala Ser His Leu Gly Arg Glu  
 125 130 135  
 Glu Thr Gln Glu Ile Arg Ala Ala Arg Arg Ser Arg Leu Arg Asp  
 140 145 150  
 Pro Ile Lys Pro Gly Met Phe Gly Tyr Gly Arg Val Pro Phe Ala  
 155 160 165  
 Leu Pro Leu His Arg Asn Arg Arg His Pro Arg Ser Pro Pro Arg  
 170 175 180  
 Ser Glu Leu Ser Leu Ile Ser Ser Arg Gly Glu Glu Pro Ile Pro  
 185 190 195  
 Ser Pro Thr Pro Arg Ala Glu Pro Phe Ser Ala Asn Gly Ser Pro  
 200 205 210  
 Gln Thr Glu Leu Pro Pro Thr Glu Leu Ser Val His Thr Pro Ser  
 215 220 225  
 Pro Gln Ala Glu Pro Leu Ser Pro Glu Thr Ala Gln Thr Glu Val  
 230 235 240  
 Ala Pro Arg Thr Arg Pro Ala Pro Leu Arg His His Pro Arg Ala  
 245 250 255  
 Gln Ala Ser Gly Thr Glu Pro Pro Ser Pro Thr His Ser Leu Gly  
 260 265 270  
 Glu Gly Gly Phe Phe Arg Ala Ser Pro Gln Pro Arg Arg Pro Ser  
 275 280 285  
 Ser Gln Gly Trp Ala Ser Pro Gln Val Ala Gly Arg Arg Pro Asp  
 290 295 300  
 Pro Phe Pro Ser Val Pro Arg Gly Arg Gly Gln Gln Gly Gln Gly  
 305 310 315  
 Pro Trp Gly Thr Gly Gly Thr Pro His Gly Pro Arg Leu Glu Pro  
 320 325 330  
 Asp Pro Gln His Pro Gly Ala Trp Leu Pro Leu Leu Ser Asn Gly  
 335 340 345  
 Pro His Ala Ser Ser Leu Trp Ser Leu Phe Ala Pro Ser Ser Pro  
 350 355 360  
 Ile Pro Arg Cys Ser Gly Glu Ser Glu Gln Leu Arg Ala Cys Ser  
 365 370 375  
 Gln Ala Pro Cys Pro Pro Glu Gln Pro Asp Pro Arg Ala Leu Gln  
 380 385 390  
 Cys Ala Ala Phe Asn Ser Gln Glu Phe Met Gly Gln Leu Tyr Gln  
 395 400 405  
 Trp Glu Pro Phe Thr Glu Val Gln Gly Ser Gln Arg Cys Glu Leu  
 410 415 420  
 Asn Cys Arg Pro Arg Gly Phe Arg Phe Tyr Val Arg His Thr Glu  
 425 430 435  
 Lys Val Gln Asp Gly Thr Leu Cys Gln Pro Gly Ala Pro Asp Ile  
 440 445 450  
 Cys Val Ala Gly Arg Cys Leu Ser Pro Gly Cys Asp Gly Ile Leu  
 455 460 465  
 Gly Ser Gly Arg Arg Pro Asp Gly Cys Gly Val Cys Gly Gly Asp  
 470 475 480

Asp Ser Thr Cys Arg Leu Val Ser Gly Asn Leu Thr Asp Arg Gly  
                   485                          490                          495  
 Gly Pro Leu Gly Tyr Gln Lys Ile Leu Trp Ile Pro Ala Gly Ala  
                   500                          505                          510  
 Leu Arg Leu Gln Ile Ala Gln Leu Arg Pro Ser Ser Asn Tyr Leu  
                   515                          520                          525  
 Ala Leu Arg Gly Pro Gly Gly Arg Ser Ile Ile Asn Gly Asn Trp  
                   530                          535                          540  
 Ala Val Asp Pro Pro Gly Ser Tyr Arg Ala Gly Gly Thr Val Phe  
                   545                          550                          555  
 Arg Tyr Asn Arg Pro Pro Arg Glu Glu Gly Lys Gly Glu Ser Leu  
                   560                          565                          570  
 Ser Ala Glu Gly Pro Thr Thr Gln Pro Val Asp Val Tyr Met Ile  
                   575                          580                          585  
 Phe Gln Glu Glu Asn Pro Gly Val Phe Tyr Gln Tyr Val Ile Ser  
                   590                          595                          600  
 Ser Pro Pro Pro Ile Leu Glu Asn Pro Thr Pro Glu Pro Pro Val  
                   605                          610                          615  
 Pro Gln Leu Gln Pro Glu Ile Leu Arg Val Glu Pro Pro Leu Ala  
                   620                          625                          630  
 Pro Ala Pro Arg Pro Ala Arg Thr Pro Gly Thr Leu Gln Arg Gln  
                   635                          640                          645  
 Val Arg Ile Pro Gln Met Pro Ala Pro His Pro Arg Thr Pro  
                   650                          655                          660  
 Leu Gly Ser Pro Ala Ala Tyr Trp Lys Arg Val Gly His Ser Ala  
                   665                          670                          675  
 Cys Ser Ala Ser Cys Gly Lys Gly Val Trp Arg Pro Ile Phe Leu  
                   680                          685                          690  
 Cys Ile Ser Arg Glu Ser Gly Glu Glu Leu Asp Glu Arg Ser Cys  
                   695                          700                          705  
 Ala Ala Gly Ala Arg Pro Pro Ala Ser Pro Glu Pro Cys His Gly  
                   710                          715                          720  
 Thr Pro Cys Pro Pro Tyr Trp Glu Ala Gly Glu Trp Thr Ser Cys  
                   725                          730                          735  
 Ser Arg Ser Cys Gly Pro Gly Thr Gln His Arg Gln Leu Gln Cys  
                   740                          745                          750  
 Arg Gln Glu Phe Gly Gly Gly Ser Ser Val Pro Pro Glu Arg  
                   755                          760                          765  
 Cys Gly His Leu Pro Arg Pro Asn Ile Thr Gln Ser Cys Gln Leu  
                   770                          775                          780  
 Arg Leu Cys Gly His Trp Glu Val Gly Ser Pro Trp Ser Gln Cys  
                   785                          790                          795  
 Ser Val Arg Cys Gly Arg Gly Gln Arg Ser Arg Gln Val Arg Cys  
                   800                          805                          810  
 Val Gly Asn Asn Gly Asp Glu Val Ser Glu Gln Glu Cys Ala Ser  
                   815                          820                          825  
 Gly Pro Pro Gln Pro Pro Ser Arg Glu Ala Cys Asp Met Gly Pro  
                   830                          835                          840  
 Cys Thr Thr Ala Trp Phe His Ser Asp Trp Ser Ser Lys Cys Ser  
                   845                          850                          855  
 Ala Glu Cys Gly Thr Gly Ile Gln Arg Arg Ser Val Val Cys Leu  
                   860                          865                          870  
 Gly Ser Gly Ala Ala Thr Arg Ala Arg Pro Gly Gly Ser Arg Ser  
                   875                          880                          885  
 Arg Asn Trp Ala Glu Leu Ser Asn Arg Lys Pro Ala Pro  
                   890                          895

<210> 13  
 <211> 631  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 3473847CD1

<400> 13

Met Phe Leu Leu Ala Trp Gly Gln Asp Pro Trp Arg Leu Pro Gly  
 1 5 10 15  
 Thr Tyr Val Val Val Leu Lys Glu Glu Thr His Leu Ser Gln Ser  
 20 25 30  
 Glu Arg Thr Ala Arg Arg Leu Gln Ala Gln Ala Ala Arg Arg Gly  
 35 40 45  
 Tyr Leu Thr Lys Ile Leu His Val Phe His Gly Leu Leu Pro Gly  
 50 55 60  
 Phe Leu Val Lys Met Ser Gly Asp Leu Leu Glu Leu Ala Leu Lys  
 65 70 75  
 Leu Pro His Val Asp Tyr Ile Glu Glu Asp Ser Ser Val Phe Ala  
 80 85 90  
 Gln Ser Ile Pro Trp Asn Leu Glu Arg Ile Thr Pro Pro Arg Tyr  
 95 100 105  
 Arg Ala Asp Glu Tyr Gln Pro Pro Asp Gly Gly Ser Leu Val Glu  
 110 115 120  
 Val Tyr Leu Leu Asp Thr Ser Ile Gln Ser Asp His Arg Glu Ile  
 125 130 135  
 Glu Gly Arg Val Met Val Thr Asp Phe Glu Asn Val Pro Glu Glu  
 140 145 150  
 Asp Gly Thr Arg Phe His Arg Gln Ala Ser Lys Cys Asp Ser His  
 155 160 165  
 Gly Thr His Leu Ala Gly Val Val Ser Gly Arg Asp Ala Gly Val  
 170 175 180  
 Ala Lys Gly Ala Ser Met Arg Ser Leu Arg Val Leu Asn Cys Gln  
 185 190 195  
 Gly Lys Gly Thr Val Ser Gly Thr Leu Ile Gly Leu Glu Phe Ile  
 200 205 210  
 Arg Lys Ser Gln Leu Val Gln Pro Val Gly Pro Leu Val Val Leu  
 215 220 225  
 Leu Pro Leu Ala Gly Tyr Ser Arg Val Leu Asn Ala Ala Cys  
 230 235 240  
 Gln Arg Leu Ala Arg Ala Gly Val Val Leu Val Thr Ala Ala Gly  
 245 250 255  
 Asn Phe Arg Asp Asp Ala Cys Leu Tyr Ser Pro Ala Ser Ala Pro  
 260 265 270  
 Glu Val Ile Thr Val Gly Ala Thr Asn Ala Gln Asp Gln Pro Val  
 275 280 285  
 Thr Leu Gly Thr Leu Gly Thr Asn Phe Gly Arg Cys Val Asp Leu  
 290 295 300  
 Phe Ala Pro Gly Glu Asp Ile Ile Gly Ala Ser Ser Asp Cys Ser  
 305 310 315  
 Thr Cys Phe Val Ser Gln Ser Gly Thr Ser Gln Ala Ala Ala His  
 320 325 330  
 Val Ala Gly Ile Ala Ala Met Met Leu Ser Ala Glu Pro Glu Leu  
 335 340 345  
 Thr Leu Ala Glu Leu Arg Gln Arg Leu Ile His Phe Ser Ala Lys  
 350 355 360  
 Asp Val Ile Asn Glu Ala Trp Phe Pro Glu Asp Gln Arg Val Leu  
 365 370 375  
 Thr Pro Asn Leu Val Ala Ala Leu Pro Pro Ser Thr His Gly Ala  
 380 385 390  
 Gly Trp Gln Leu Phe Cys Arg Thr Val Trp Ser Ala His Ser Gly  
 395 400 405  
 Pro Thr Arg Met Ala Thr Ala Ile Ala Arg Cys Ala Pro Asp Glu  
 410 415 420  
 Glu Leu Leu Ser Cys Ser Ser Phe Ser Arg Ser Gly Lys Arg Arg  
 425 430 435  
 Gly Glu Arg Met Glu Ala Gln Gly Gly Lys Leu Val Cys Arg Ala  
 440 445 450  
 His Asn Ala Phe Gly Gly Glu Gly Val Tyr Ala Ile Ala Arg Cys  
 455 460 465  
 Cys Leu Leu Pro Gln Ala Asn Cys Ser Val His Thr Ala Pro Pro  
 470 475 480  
 Ala Glu Ala Ser Met Gly Thr Arg Val His Cys His Gln Gln Gly  
 485 490 495  
 His Val Leu Thr Gly Cys Ser Ser His Trp Glu Val Glu Asp Leu

	500	505	510
Gly Thr His Lys Pro Pro Val Leu Arg	Pro Arg Gly Gln Pro	Asn	
515	520	525	
Gln Cys Val Gly His Arg Glu Ala Ser	Ile His Ala Ser Cys	Cys	
530	535	540	
His Ala Pro Gly Leu Glu Cys Lys Val	Lys Glu His Gly Ile	Pro	
545	550	555	
Ala Pro Gln Glu Gln Val Thr Val Ala	Cys Glu Glu Gly Trp	Thr	
560	565	570	
Leu Thr Gly Cys Ser Ala Leu Pro Gly	Thr Ser His Val Leu	Gly	
575	580	585	
Ala Tyr Ala Val Asp Asn Thr Cys Val	Val Arg Ser Arg Asp	Val	
590	595	600	
Ser Thr Thr Gly Ser Thr Ser Glu Glu	Ala Val Thr Ala Val	Ala	
605	610	615	
Ile Cys Cys Arg Ser Arg His Leu Ala	Gln Ala Ser Gln Glu	Leu	
620	625	630	
Gln			

<210> 14  
<211> 470  
<212> PRT  
<213> Homo sapiens

<220>  
<221> misc\_feature  
<223> Incyte ID No: 3750004CD1

<400> 14  
Met Arg His Arg Thr Asp Leu Gly Gln Asn Leu Leu Leu Phe Leu  
1 5 10 15  
Trp Ala Leu Leu Asn Cys Gly Leu Gly Val Ser Ala Gln Gly Pro  
20 25 30  
Gly Glu Trp Thr Pro Trp Val Ser Trp Thr Arg Cys Ser Ser Ser  
35 40 45  
Cys Gly Arg Gly Val Ser Val Arg Ser Arg Arg Cys Leu Arg Leu  
50 55 60  
Pro Gly Glu Pro Cys Trp Gly Asp Ser His Glu Tyr Arg Leu  
65 70 75  
Cys Gln Leu Pro Asp Cys Pro Pro Gly Ala Val Pro Phe Arg Asp  
80 85 90  
Leu Gln Cys Ala Leu Tyr Asn Gly Arg Pro Val Leu Gly Thr Gln  
95 100 105  
Lys Thr Tyr Gln Trp Val Pro Phe His Gly Ala Pro Asn Gln Cys  
110 115 120  
Asp Leu Asn Cys Leu Ala Glu Gly His Ala Phe Tyr His Ser Phe  
125 130 135  
Gly Arg Val Leu Asp Gly Thr Ala Cys Ser Pro Gly Ala Gln Gly  
140 145 150  
Val Cys Val Ala Gly Arg Cys Leu Ser Ala Gly Cys Asp Gly Leu  
155 160 165  
Leu Gly Ser Gly Ala Leu Glu Asp Arg Cys Gly Arg Cys Gly Gly  
170 175 180  
Ala Asn Asp Ser Cys Leu Phe Val Gln Arg Val Phe Arg Asp Ala  
185 190 195  
Gly Ala Phe Ala Gly Tyr Trp Asn Val Thr Leu Ile Pro Glu Gly  
200 205 210  
Ala Arg His Ile Arg Val Glu His Arg Ser Arg Asn His Leu Gly  
215 220 225  
Ile Leu Gly Ser Leu Met Gly Gly Asp Gly Arg Tyr Val Leu Asn  
230 235 240  
Gly His Trp Val Val Ser Pro Pro Gly Thr Tyr Glu Ala Ala Gly  
245 250 255  
Thr His Val Val Tyr Thr Arg Asp Thr Gly Pro Gln Glu Thr Leu  
260 265 270  
Gln Ala Ala Gly Pro Thr Ser His Asp Leu Leu Gln Val Leu

Leu Gln Glu Pro Asn Pro Gly Ile Glu	275	Phe Glu Phe Trp Leu	280	Pro
290	295	300		
Arg Glu Arg Tyr Ser Pro Phe Gln Ala	305	Arg Val Gln Ala Leu	310	Gly
315				
Trp Pro Leu Arg Gln Pro Gln Pro Arg	320	Gly Val Glu Pro Gln	325	Pro
330	335	330		
Pro Ala Ala Pro Ala Val Thr Pro Ala	335	Gln Thr Pro Thr Leu	340	Ala
345				
Pro Asp Pro Cys Pro Pro Cys Pro Asp	350	Thr Arg Gly Arg Ala	355	His
360				
Arg Leu Leu His Tyr Cys Gly Ser Asp	365	Phe Val Phe Gln Ala	370	Arg
375				
Val Leu Gly His His His Gln Ala Gln	380	Glu Thr Arg Tyr Glu	385	Val
390				
Arg Ile Gln Leu Val Tyr Lys Asn Arg	395	Ser Pro Leu Arg Ala	400	Arg
405				
Glu Tyr Val Trp Ala Pro Gly His Cys	410	Pro Cys Pro Met Leu	415	Ala
420				
Pro His Arg Asp Tyr Leu Met Ala Val	425	Gln Arg Leu Val Ser	430	Pro
435				
Asp Gly Thr Gln Asp Gln Leu Leu Leu	440	Pro His Ala Gly Tyr	445	Ala
450				
Arg Pro Trp Ser Pro Ala Glu Asp Ser	455	Arg Ile Arg Leu Thr	460	Ala
465				
Arg Arg Cys Pro Gly	470			

<210> 15  
<211> 110  
<212> PRT  
<213> Homo sapiens

<220>  
<221> misc\_feature  
<223> Incyte ID No: 4904126CD1

Met Ala Asp Lys Val Leu Lys Glu Lys Arg Lys Gln Phe Ile Arg	1	5	10	15
Ser Val Gly Glu Gly Thr Ile Asn Gly Leu Leu Gly Glu Leu Leu	20	25	30	
Glu Thr Arg Val Leu Ser Gln Glu Glu Ile Glu Ile Val Lys Cys	35	40	45	
Glu Asn Ala Thr Val Met Asp Lys Ala Arg Ala Leu Leu Asp Ser	50	55	60	
Val Ile Arg Lys Gly Ala Pro Ala Cys Gln Ile Cys Ile Thr Tyr	65	70	75	
Ile Cys Glu Glu Asp Ser His Leu Ala Gly Thr Leu Gly Leu Ser	80	85	90	
Ala Gly Pro Thr Ser Gly Asn His Leu Thr Thr Gln Asp Ser Gln	95	100	105	
Ile Val Leu Pro Ser	110			

<210> 16  
<211> 879  
<212> PRT  
<213> Homo sapiens

<220>  
<221> misc\_feature  
<223> Incyte ID No: 71268415CD1

Met Ser Leu Phe Ile Phe Cys Arg Gln Leu Phe Ala Pro Ser Tyr	1	5	10	15
---	---	---	----	----

Thr Glu Thr His Tyr Thr Ser Ser Gly Asn Pro Gln Thr Thr Thr  
                   20                  25                  30  
 Arg Lys Leu Glu Asp His Cys Phe Tyr His Gly Thr Val Arg Glu  
                   35                  40                  45  
 Thr Glu Leu Ser Ser Val Thr Leu Ser Thr Cys Arg Gly Ile Arg  
                   50                  55                  60  
 Gly Leu Ile Thr Val Ser Ser Asn Leu Ser Tyr Val Ile Glu Pro  
                   65                  70                  75  
 Leu Pro Asp Ser Lys Gly Gln His Leu Ile Tyr Arg Ser Glu His  
                   80                  85                  90  
 Leu Lys Pro Pro Pro Gly Asn Cys Gly Phe Glu His Ser Lys Pro  
                   95                  100                105  
 Thr Thr Arg Asp Trp Ala Leu Gln Phe Thr Gln Gln Thr Lys Lys  
                   110                115                120  
 Arg Pro Arg Arg Met Lys Arg Glu Asp Leu Asn Ser Met Lys Tyr  
                   125                130                135  
 Val Glu Leu Tyr Leu Val Ala Asp Tyr Leu Glu Phe Gln Lys Asn  
                   140                145                150  
 Arg Arg Asp Gln Asp Ala Thr Lys His Lys Leu Ile Glu Ile Ala  
                   155                160                165  
 Asn Tyr Val Asp Lys Phe Tyr Arg Ser Leu Asn Ile Arg Ile Ala  
                   170                175                180  
 Leu Val Gly Leu Glu Val Trp Thr His Gly Asn Met Cys Glu Val  
                   185                190                195  
 Ser Glu Asn Pro Tyr Ser Thr Leu Trp Ser Phe Leu Ser Trp Arg  
                   200                205                210  
 Arg Lys Leu Leu Ala Gln Lys Tyr His Asp Asn Ala Gln Leu Ile  
                   215                220                225  
 Thr Gly-Met-Ser Phe His Gly Thr Thr Ile Gly Leu Ala Pro Leu  
                   230                235                240  
 -Met-Ala-Met-Cys Ser Val Tyr Gln Ser Gly Gly Val Asn Met Asp  
                   245                250                255  
 -His-Ser-Glu-Asn-Ala-Ile-Gly Val-Ala Ala Thr Met Ala His Glu  
                   260                265                270  
 -Met-Gly-His-Asn-Phe Gly-Met-Thr-His Asp Ser Ala Asp Cys Cys  
                   275                280                285  
 -Ser-Ala-Ser-Ala-Ala Asp-Gly Gly Cys Ile Met Ala Ala Ala Thr  
                   290                295                300  
 Gly-His-Pro-Phe-Lys-Val-Phe-Asn Gly Cys Asn Arg Arg Glu  
                   305                310                315  
 Leu Asp-Arg-Tyr-Leu-Gln Ser Gly Gly Met Cys Leu Ser Asn  
                   320                325                330  
 Met-Pro-Asp-Thr-Arg Met Leu Tyr Gly Gly Arg Arg Cys Gly Asn  
                   335                340                345  
 Gly-Tyr-Leu-Glu-Asp Gly Glu Glu Cys Asp Cys Gly Glu Glu Glu  
                   350                355                360  
 Glu Cys-Asn-Asn-Pro-Cys Cys Asn Ala Ser Asn Cys Thr Leu Arg  
                   365                370                375  
 Pro Gly Ala-Glu Cys Ala His Gly Ser Cys Cys His Gln Cys Lys  
                   380                385                390  
 Leu Leu Ala Pro Gly Thr Leu Cys Arg Glu Gln Ala Arg Gln Cys  
                   395                400                405  
 Asp Leu Pro Glu Phe Cys Thr Gly Lys Ser Pro His Cys Pro Thr  
                   410                415                420  
 Asn Phe Tyr Gln Met Asp Gly Thr Pro Cys Glu Gly Gln Ala  
                   425                430                435  
 Tyr Cys Tyr Asn Gly Met Cys Leu Thr Tyr Gln Glu Gln Cys Gln  
                   440                445                450  
 Gln Leu Trp Gly Pro Gly Ala Arg Pro Ala Pro Asp Leu Cys Phe  
                   455                460                465  
 Glu Lys Val Asn Val Ala Gly Asp Thr Phe Gly Asn Cys Gly Lys  
                   470                475                480  
 Asp Met Asn Gly Glu His Arg Lys Cys Asn Met Arg Asp Ala Lys  
                   485                490                495  
 Cys Gly Lys Ile Gln Cys Gln Ser Ser Glu Ala Arg Pro Leu Glu  
                   500                505                510  
 Ser Asn Ala Val Pro Ile Asp Thr Thr Ile Ile Met Asn Gly Arg

Gln	Ile	Gln	Cys	Arg	Gly	Thr	His	Val	Tyr	Arg	Gly	Pro	Glu	Glu
515														525
				530				535						540
Glu	Gly	Asp	Met	Leu	Asp	Pro	Gly	Leu	Val	Met	Thr	Gly	Thr	Lys
														545
Cys	Gly	Tyr	Asn	His	Ile	Cys	Phe	Glu	Gly	Gln	Cys	Arg	Asn	Thr
														550
				560				565						555
Ser	Phe	Phe	Glu	Thr	Glu	Gly	Cys	Gly	Lys	Lys	Cys	Asn	Gly	His
														560
				575				580						570
Gly	Val	Cys	Asn	Asn	Asn	Gln	Asn	Cys	His	Cys	Leu	Pro	Gly	Trp
														585
Ala	Pro	Pro	Phe	Cys	Asn	Thr	Pro	Gly	His	Gly	Gly	Ser	Ile	Asp
														590
				605				610						600
Ser	Gly	Pro	Met	Pro	Pro	Glu	Ser	Val	Gly	Pro	Val	Val	Ala	Gly
														615
				620				625						630
Val	Leu	Val	Ala	Ile	Leu	Val	Leu	Ala	Val	Leu	Met	Leu	Met	Tyr
														635
Tyr	Cys	Cys	Arg	Gln	Asn	Asn	Lys	Leu	Gly	Gln	Leu	Lys	Pro	Ser
														640
				650				655						645
Ala	Leu	Pro	Ser	Lys	Leu	Arg	Gln	Gln	Phe	Ser	Cys	Pro	Phe	Arg
														655
				665				670						660
Val	Ser	Gln	Asn	Ser	Gly	Thr	Gly	His	Ala	Asn	Pro	Thr	Phe	Lys
														675
				680				685						690
Leu	Gln	Thr	Pro	Gln	Gly	Lys	Arg	Lys	Val	Ile	Asn	Thr	Pro	Glu
														695
Ile	Leu	Arg	Lys	Pro	Ser	Gln	Pro	Pro	Pro	Arg	Pro	Pro	Pro	Asp
														700
				710				715						705
Tyr	Leu	Arg	Gly	Gly	Ser	Pro	Pro	Ala	Pro	Leu	Pro	Ala	His	Leu
														725
				725				730						735
Ser	Arg	Ala	Ala	Arg	Asn	Ser	Pro	Gly	Pro	Gly	Ser	Gln	Ile	Glu
														740
				740				745						750
Arg	Thr	Glu	Ser	Ser	Arg	Arg	Pro	Pro	Pro	Ser	Arg	Pro	Ile	Pro
														755
				755				760						765
Pro	Ala	Pro	Asn	Cys	Ile	Val	Ser	Gln	Asp	Phe	Ser	Arg	Pro	Arg
														770
				770				775						780
Pro	Pro	Gln	Lys	Ala	Leu	Pro	Ala	Asn	Pro	Val	Pro	Gly	Arg	Arg
														785
				785				790						795
Ser	Leu	Pro	Arg	Pro	Gly	Gly	Ala	Ser	Pro	Leu	Arg	Pro	Gly	
														800
				800				805						810
Ala	Gly	Pro	Gln	Gln	Ser	Arg	Pro	Leu	Ala	Ala	Leu	Ala	Pro	Lys
														815
				815				820						825
Val	Ser	Pro	Arg	Glu	Ala	Leu	Lys	Val	Lys	Ala	Gly	Thr	Arg	Gly
														830
				830				835						840
Leu	Gln	Gly	Gly	Arg	Cys	Arg	Val	Glu	Lys	Thr	Lys	Gln	Phe	Met
														845
				845				850						855
Leu	Leu	Val	Val	Trp	Thr	Glu	Leu	Pro	Glu	Gln	Lys	Pro	Arg	Ala
														860
Lys	His	Ser	Cys	Phe	Leu	Val	Pro	Ala	865					870
				875										

<210> 17  
<211> 850  
<212> PRT  
<213> Homo sapiens

<220>  
<221> misc\_feature  
<223> Incyte ID No: 7473301CD1

<400> 17  
Met Asp Lys Glu Asn Ser Asp Val Ser Ala Ala Pro Ala Asp Leu  
1 5 10 15  
Lys Ile Ser Asn Ile Ser Val Gln Val Val Ser Ala Gln Lys Lys  
20 25 30  
Leu Pro Val Arg Arg Pro Pro Leu Pro Gly Arg Arg Leu Pro Leu  
35 40 45  
Pro Gly Arg Arg Pro Pro Gln Arg Pro Ile Gly Lys Ala Lys Pro

50	55	60
Lys Lys Gln Ser Lys	Lys Lys Val Pro Phe Trp Asn Val Gln Asn	
65	70	75
Lys Ile Ile Leu Phe Thr Val Phe Leu Phe	Ile Leu Ala Val Ile	
80	85	90
Ala Trp Thr Leu Leu Trp Leu Tyr Ile Ser	Lys Thr Glu Ser Lys	
95	100	105
Asp Ala Phe Tyr Phe Ala Gly Met Phe Arg	Ile Thr Asn Ile Glu	
110	115	120
Phe Leu Pro Glu Tyr Arg Gln Lys Glu Ser	Arg Glu Phe Leu Ser	
125	130	135
Val Ser Arg Thr Val Gln Gln Val Ile Asn	Leu Val Tyr Thr	
140	145	150
Ser Ala Phe Ser Lys Phe Tyr Glu Gln Ser	Val Val Ala Asp Val	
155	160	165
Ser Ser Asn Asn Lys Gly Gly Leu Leu Val	His Phe Trp Ile Val	
170	175	180
Phe Val Met Pro Arg Ala Lys Gly His	Ile Phe Cys Glu Asp Cys	
185	190	195
Val Ala Ala Ile Leu Lys Asp Ser Ile	Gln Thr Ser Ile Ile Asn	
200	205	210
Arg Thr Ser Val Gly Ser Leu Gln Gly	Leu Ala Val Asp Met Asp	
215	220	225
Ser Val Val Leu Asn Gly Asp Cys Trp	Ser Phe Leu Lys Lys	
230	235	240
Lys Arg Lys Glu Asn Gly Ala Val Ser	Thr Asp Lys Gly Cys Ser	
245	250	255
Gln Tyr Phe Tyr Ala Glu His Leu Ser	Leu His Tyr Pro Leu Glu	
260	265	270
Ile Ser Ala Ala Ser Gly Arg Leu Met	Cys His Phe Lys Leu Val	
275	280	285
Ala Ile Val Gly Tyr Leu Ile Arg Leu Ser	Ile Lys Ser Ile Gln	
290	295	300
Ile Glu Ala Asp Asn Cys Val Thr Asp	Ser Leu Thr Ile Tyr Asp	
305	310	315
Ser Leu Leu Pro Ile Arg Ser Ser Ile	Leu Tyr Arg Ile Cys Glu	
320	325	330
Pro Thr Arg Thr Leu Met Ser Phe Val	Ser Thr Asn Asn Leu Met	
335	340	345
Leu Val Thr Phe Lys Ser Pro His Ile Arg	Arg Leu Ser Gly Ile	
350	355	360
Arg Ala Tyr Phe Glu Val Ile Pro Glu	Gln Lys Cys Glu Asn Thr	
365	370	375
Val Leu Val Lys Asp Ile Thr Gly Phe	Glu Gly Lys Ile Ser Ser	
380	385	390
Pro Tyr Tyr Pro Ser Tyr Tyr Pro Pro	Lys Cys Lys Cys Thr Trp	
395	400	405
Lys Phe Gln Thr Ser Leu Ser Thr Leu Gly	Ile Ala Leu Lys Phe	
410	415	420
Tyr Asn Tyr Ser Ile Thr Lys Lys Ser	Met Lys Gly Cys Glu His	
425	430	435
Gly Trp Trp Glu Ile Tyr Glu His Met	Tyr Cys Gly Ser Tyr Met	
440	445	450
Asp His Gln Thr Ile Phe Arg Val Pro	Ser Pro Leu Val His Ile	
455	460	465
Gln Leu Gln Cys Ser Ser Arg Leu Ser	Gly Lys Pro Leu Leu Ala	
470	475	480
Glu Tyr Gly Ser Tyr Asn Ile Ser Gln	Pro Cys Pro Val Gly Ser	
485	490	495
Phe Arg Cys Ser Ser Gly Leu Cys Val	Pro Gln Ala Gln Arg Gly	
500	505	510
Asp Gly Val Asn Asp Cys Phe Asp Glu	Ser Asp Glu Leu Phe Cys	
515	520	525
Val Ser Pro Gln Pro Ala Cys Asn Thr	Ser Ser Phe Arg Gln His	
530	535	540
Gly Pro Leu Ile Cys Asp Gly Phe Arg	Asp Cys Glu Asn Gly Arg	
545	550	555

Asp Glu Gln Asn Cys Thr Gln Ser Ile Pro Cys Asn Asn Arg Thr  
     560                 565                 570  
 Phe Lys Cys Gly Asn Asp Ile Cys Phe Arg Lys Gln Asn Ala Lys  
     575                 580                 585  
 Cys Asp Gly Thr Val Asp Cys Pro Asp Gly Ser Asp Glu Glu Gly  
     590                 595                 600  
 Cys Thr Cys Ser Arg Ser Ser Ser Ala Leu His Arg Ile Ile Gly  
     605                 610                 615  
 Gly Thr Asp Thr Leu Glu Gly Gly Trp Pro Trp Gln Val Ser Leu  
     620                 625                 630  
 His Phe Val Gly Ser Ala Tyr Cys Gly Ala Ser Val Ile Ser Arg  
     635                 640                 645  
 Glu Trp Leu Leu Ser Ala Ala His Cys Phe His Gly Asn Arg Leu  
     650                 655                 660  
 Ser Asp Pro Thr Pro Trp Thr Ala His Leu Gly Met Tyr Val Gln  
     665                 670                 675  
 Gly Asn Ala Lys Phe Val Ser Pro Val Arg Arg Ile Val Val His  
     680                 685                 690  
 Glu Tyr Tyr Asn Ser Gln Thr Phe Asp Tyr Asp Ile Ala Leu Leu  
     695                 700                 705  
 Gln Leu Ser Ile Ala Trp Pro Glu Thr Leu Lys Gln Leu Ile Gln  
     710                 715                 720  
 Pro Ile Cys Ile Pro Pro Thr Gly Gln Arg Val Arg Ser Gly Glu  
     725                 730                 735  
 Lys Cys Trp Val Thr Gly Trp Gly Arg Arg His Glu Ala Asp Asn  
     740                 745                 750  
 Lys Gly Ser Leu Val Leu Gln Gln Ala Glu Val Glu Leu Ile Asp  
     755                 760                 765  
 Gln Thr Leu Cys Val Ser Thr Tyr Gly Ile Ile Thr Ser Arg Met  
     770                 775                 780  
 Leu Cys Ala Gly Ile Met Ser Gly Lys Arg Asp Ala Cys Lys Gly  
     785                 790                 795  
 Asp Ser Gly Gly Pro Leu Ser Cys Arg Arg Lys Ser Asp Gly Lys  
     800                 805                 810  
 Trp Ile Leu Thr Gly Ile Val Ser Trp Gly His Gly Cys Gly Arg  
     815                 820                 825  
 Pro Asn Phe Pro Gly Val Tyr Thr Arg Val Ser Asn Phe Val Pro  
     830                 835                 840  
 Trp Ile His Lys Tyr Val Pro Ser Leu Leu  
     845                 850

<210> 18  
 <211> 254  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 7473308CD1

<400> 18

Met	Gln	Asp	His	Arg	Lys	Gly	Lys	Ala	Ala	Val	Gly	Val	Ser	Phe
1				5						10				15
Asp	Asp	Asp	Asp	Lys	Ile	Val	Gly	Gly	Tyr	Asn	Cys	Glu	Glu	Asn
				20					25					30
Ser	Val	Pro	Tyr	Gln	Val	Ser	Leu	Asn	Ser	Gly	Tyr	His	Phe	Cys
				35					40					45
Val	Gly	Ser	Leu	Asn	Arg	Glu	Tyr	Cys	Ile	Gln	Val	Arg	Leu	Gly
				50					55					60
Glu	His	Asn	Ile	Glu	Val	Leu	Glu	Gly	Asn	Glu	Gln	Phe	Ile	Tyr
				65					70					75
Ala	Val	Lys	Ile	Ile	Arg	His	Pro	Lys	Tyr	Asn	Ser	Trp	Thr	Leu
				80					85					90
Asp	Asn	Asp	Ile	Leu	Leu	Ile	Lys	Leu	Ser	Thr	Pro	Ala	Ile	Ile
				95					100					105
Asn	Ala	His	Val	Ser	Thr	Ile	Ser	Leu	Pro	Thr	Thr	Pro	Pro	Ala
				110					115					120

Ala Gly Thr Glu Cys Leu Ile Ser Gly Trp Gly Asn Thr Leu Ser  
                   125                  130                  135  
 Ser Gly Ala Asp Tyr Pro Asp Glu Leu Gln Cys Leu Asp Ala Pro  
                   140                  145                  150  
 Val Leu Ser Gln Ala Glu Tyr Glu Ala Ser Tyr Pro Gly Lys Ile  
                   155                  160                  165  
 Thr Asn Asn Val Phe Cys Val Gly Phe Leu Glu Gly Gly Lys Asp  
                   170                  175                  180  
 Ser Cys Gln Ile Ile Pro Ile Lys Val Gln Gln Leu Val Thr Ser  
                   185                  190                  195  
 Ser Gln Glu Thr Asp Ile Arg Ile Pro Met Ala Leu Gln Thr Ala  
                   200                  205                  210  
 Ala Ser Thr Ser Tyr Leu Gly Pro Leu Asp Ser Leu His Arg Lys  
                   215                  220                  225  
 Val Ser His Pro Thr Glu Lys Arg Cys Gln Gln Lys Gln Gly Met  
                   230                  235                  240  
 Lys Ile Thr Asp Asn His Gly Ile Thr Ser Lys Trp Ser Val  
                   245                  250

<210> 19  
 <211> 568  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 7478021CD1

<400> 19

Met	Leu	Ala	Ala	Ser	Ile	Phe	Arg	Pro	Thr	Leu	Leu	Leu	Cys	Trp
1														
														15
Leu	Ala	Ala	Pro	Trp	Pro	Thr	Gln	Pro	Glu	Ser	Leu	Phe	His	Ser
														30
Arg	Asp	Arg	Ser	Asp	Leu	Glu	Pro	Ser	Pro	Leu	Arg	Gln	Ala	Lys
														45
Pro	Ile	Ala	Asp	Leu	His	Ala	Ala	Gln	Arg	Phe	Leu	Ser	Arg	Tyr
														60
Gly	Trp	Ser	Gly	Val	Trp	Ala	Ala	Trp	Gly	Pro	Ser	Pro	Glu	Gly
														75
+	Pro	Pro	Glu	Thr	Pro	Lys	Gly	Ala	Ala	Leu	Ala	Glu	Ala	Val
														90
Arg	Phe	Gln	Arg	Ala	Asn	Ala	Leu	Pro	Ala	Ser	Gly	Glu	Leu	Asp
														105
Ala	Ala	Thr	Leu	Ala	Ala	Met	Asn	Arg	Pro	Arg	Cys	Gly	Val	Pro
														120
Asp	Met	Arg	Pro	Pro	Pro	Pro	Ser	Ala	Pro	Pro	Ser	Pro	Pro	Gly
														135
Pro	Pro	Pro	Arg	Ala	Arg	Ser	Arg	Arg	Ser	Pro	Arg	Ala	Pro	Leu
														150
Ser	Leu	Ser	Arg	Arg	Gly	Trp	Gln	Pro	Arg	Gly	Tyr	Pro	Asp	Gly
														165
Gly	Ala	Ala	Gln	Ala	Phe	Ser	Lys	Arg	Thr	Leu	Ser	Trp	Arg	Leu
														180
Leu	Gly	Glu	Ala	Leu	Ser	Ser	Gln	Leu	Ser	Val	Ala	Asp	Gln	Arg
														195
Arg	Ile	Glu	Ala	Leu	Ala	Phe	Arg	Met	Trp	Ser	Glu	Val	Thr	Pro
														210
Leu	Asp	Phe	Arg	Glu	Asp	Leu	Ala	Ala	Pro	Gly	Ala	Ala	Val	Asp
														225
Ile	Lys	Leu	Gly	Phe	Gly	Arg	Arg	His	Leu	Gly	Cys	Pro	Arg	Ala
														240
Phe	Asp	Gly	Ser	Gly	Gln	Glu	Phe	Ala	His	Ala	Trp	Arg	Leu	Gly
														255
Asp	Ile	His	Phe	Asp	Asp	Asp	Glu	His	Phe	Thr	Pro	Pro	Thr	Ser
														270
Asp	Thr	Gly	Ile	Ser	Leu	Leu	Lys	Val	Ala	Val	His	Glu	Ile	Gly
														285

His Val Leu Gly Leu Pro His Thr Tyr Arg Thr Gly Ser Ile Met  
 290 295 300  
 Gln Pro Asn Tyr Ile Pro Gln Glu Pro Ala Phe Glu Leu Asp Trp  
 305 310 315  
 Ser Asp Arg Lys Ala Ile Gln Lys Leu Tyr Gly Ser Cys Glu Gly  
 320 325 330  
 Ser Phe Asp Thr Ala Phe Asp Trp Ile Arg Lys Glu Arg Asn Gln  
 335 340 345  
 Tyr Gly Glu Val Met Val Arg Phe Ser Thr Tyr Phe Phe Arg Asn  
 350 355 360  
 Ser Trp Tyr Trp Leu Tyr Glu Asn Arg Asn Asn Arg Thr Arg Tyr  
 365 370 375  
 Gly Asp Pro Ile Gln Ile Leu Thr Gly Trp Pro Gly Ile Pro Thr  
 380 385 390  
 His Asn Ile Asp Ala Phe Val His Ile Trp Thr Trp Lys Arg Asp  
 395 400 405  
 Glu Arg Tyr Phe Phe Gln Gly Asn Gln Tyr Trp Arg Tyr Asp Ser  
 410 415 420  
 Asp Lys Asp Gln Ala Leu Thr Glu Asp Glu Gln Gly Lys Ser Tyr  
 425 430 435  
 Pro Lys Leu Ile Ser Glu Gly Phe Pro Gly Ile Pro Ser Pro Leu  
 440 445 450  
 Asp Thr Ala Phe Tyr Asp Arg Arg Gln Lys Leu Ile Tyr Phe Phe  
 455 460 465  
 Lys Glu Ser Leu Val Phe Ala Phe Asp Val Asn Arg Asn Arg Val  
 470 475 480  
 Leu Asn Ser Tyr Pro Lys Arg Ile Thr Glu Val Phe Pro Ala Val  
 485 490 495  
 Ile Pro Gln Asn His Pro Phe Arg Asn Ile Asp Ser Ala Tyr Tyr  
 500 505 510  
 Ser Tyr Ala Tyr Asn Ser Ile Phe Phe Phe Lys Gly Asn Ala Tyr  
 515 520 525  
 Trp Lys Val Val Asn Asp Lys Asp Lys Gln Gln Asn Ser Trp Leu  
 530 535 540  
 Pro Ala Asn Gly Leu Phe Pro Lys Lys Phe Ile Ser Glu Lys Trp  
 545 550 555  
 Phe Asp Val Cys Asp Val His Ile Ser Thr Leu Asn Met  
 560 565

<210> 20  
 <211> 306  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 4333459CD1

<400> 20

Met	Ser	Leu	Lys	Met	Leu	Ile	Ser	Arg	Asn	Lys	Leu	Ile	Leu	Leu
1				5						10				15
Leu	Gly	Ile	Val	Phe	Phe	Glu	Arg	Gly	Lys	Ser	Ala	Thr	Leu	Ser
										20			25	30
Leu	Pro	Lys	Ala	Pro	Ser	Cys	Gly	Gln	Ser	Leu	Val	Lys	Val	Gln
									35			40		45
Pro	Trp	Asn	Tyr	Phe	Asn	Ile	Phe	Ser	Arg	Ile	Leu	Gly	Gly	Ser
						50			55			60		
Gln	Val	Glu	Lys	Gly	Ser	Tyr	Pro	Trp	Gln	Val	Ser	Leu	Lys	Gln
						65			70			75		
Arg	Gln	Lys	His	Ile	Cys	Gly	Gly	Ser	Ile	Val	Ser	Pro	Gln	Trp
						80			85			90		
Val	Ile	Thr	Ala	Ala	His	Cys	Ile	Ala	Asn	Arg	Asn	Ile	Val	Ser
						95			100			105		
Thr	Leu	Asn	Val	Thr	Ala	Gly	Glu	Tyr	Asp	Leu	Ser	Gln	Thr	Asp
						110			115			120		
Pro	Gly	Glu	Gln	Thr	Leu	Thr	Ile	Glu	Thr	Val	Ile	Ile	His	Pro
						125			130			135		

His Phe Ser Thr Lys Lys Pro Met Asp Tyr Asp Ile Ala Leu Leu  
           140                   145                   150  
 Lys Met Ala Gly Ala Phe Gln Phe Gly His Phe Val Gly Pro Ile  
           155                   160                   165  
 Cys Leu Pro Glu Leu Arg Glu Gln Phe Glu Ala Gly Phe Ile Cys  
           170                   175                   180  
 Thr Thr Ala Gly Trp Gly Arg Leu Thr Glu Gly Gly Val Leu Ser  
           185                   190                   195  
 Gln Val Leu Gln Glu Val Asn Leu Pro Ile Leu Thr Trp Glu Glu  
           200                   205                   210  
 Cys Val Ala Ala Leu Leu Thr Leu Lys Arg Pro Ile Ser Gly Lys  
           215                   220                   225  
 Thr Phe Leu Cys Thr Gly Phe Pro Asp Gly Gly Arg Asp Ala Cys  
           230                   235                   240  
 Gln Gly Asp Ser Gly Gly Ser Leu Met Cys Arg Asn Lys Lys Gly  
           245                   250                   255  
 Ala Trp Thr Leu Ala Gly Val Thr Ser Trp Gly Leu Gly Cys Gly  
           260                   265                   270  
 Arg Gly Trp Arg Asn Asn Val Arg Lys Ser Asp Gln Gly Ser Pro  
           275                   280                   285  
 Gly Ile Phe Thr Asp Ile Ser Lys Val Leu Ser Trp Ile His Glu  
           290                   295                   300  
 His Ile Gln Thr Gly Asn  
           305

&lt;210&gt; 21

&lt;211&gt; 953

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No.: 6817347CD1

&lt;400&gt; 21

Met Thr Leu Leu Ala Pro Trp Tyr Thr Gly Pro Met Ile Pro Met  
   1               5               10               15  
 Asp Val Asn Glu Pro Ser Ser Val Thr Thr Ala Pro Thr Leu Ser  
   20               25               30  
 Ser Ser Leu Gln His Ile Ser Ser Phe Leu Ala Thr Gly Lys Lys  
   35               40               45  
 Leu Ser Leu His Phe Gly His Pro Arg Glu Cys Glu Val Thr Arg  
   50               55               60  
 Ile Asp Asp Lys Asn Arg Arg Gly Leu Glu Asp Ser Glu Pro Gly  
   65               70               75  
 Ala Lys Leu Phe Asn Asn Asp Gly Val Cys Cys Cys Leu Gln Lys  
   80               85               90  
 Arg Gly Pro Val Asn Ile Thr Ser Val Cys Val Ser Pro Arg Thr  
   95               100              105  
 Leu Gln Ile Ser Val Phe Val Leu Ser Glu Lys Tyr Glu Gly Ile  
   110              115              120  
 Val Lys Phe Glu Ser Asp Glu Leu Pro Phe Gly Val Ile Gly Ser  
   125              130              135  
 Asn Ile Gly Asp Ala His Phe Gln Glu Phe Arg Ala Gly Ile Ser  
   140              145              150  
 Trp Lys Pro Val Val Asp Pro Asp Asp Pro Ile Pro Gln Phe Pro  
   155              160              165  
 Asp Cys Cys Ser Ser Ser Ser Arg Ile Pro Ser Val Ser Val  
   170              175              180  
 Leu Val Ala Val Pro Leu Val Ala Gly His Lys Gly Gln Ala Phe  
   185              190              195  
 Ile Glu Arg Met Leu Gly Cys Phe Lys Glu Leu Lys Gln Glu Leu  
   200              205              210  
 Thr Gln Glu Gly Pro Gly Gly His Pro Arg Ser Ala Trp Pro  
   215              220              225  
 Pro Arg Arg His Ala Gln Trp Pro Pro Glu Pro Cys Glu Gln Gly  
   230              235              240

Glu Glu Pro Pro Pro Val Glu Ala Glu Glu Val Glu Glu Ala Glu  
 245 250 255  
 Thr Ala Glu Lys Ala Glu Arg Lys Val Glu Ala Glu Ala Lys Val  
 260 265 270  
 Glu Gly Lys Ala Glu Ala Ala Gly Lys Ala Glu Ala Ala Gly Lys  
 275 280 285  
 Val Asp Ala Thr Glu Lys Val Glu Thr Ala Gly Lys Val Asp Ala  
 290 295 300  
 Ala Gly Lys Val Glu Thr Ala Glu Gly Pro Gly Arg Arg Ala Glu  
 305 310 315  
 Leu Lys Leu Glu Pro Glu Pro Glu Pro Val Arg Glu Ala Glu Gln  
 320 325 330  
 Glu Pro Lys Gln Glu Leu Glu Asp Glu Asn Pro Ala Arg Ser Gly  
 335 340 345  
 Gly Gly Gly Asn Ser Asp Glu Val Pro Pro Pro Thr Leu Pro Ser  
 350 355 360  
 Asp Pro Pro Arg Pro Pro Asp Pro Ser Pro Arg Arg Ser Arg Ala  
 365 370 375  
 Pro Arg Arg Arg Pro Arg Pro Arg Pro Gln Thr Arg Leu Arg Thr  
 380 385 390  
 Pro Pro Gln Pro Arg Pro Arg Pro Pro Pro Arg Pro Arg Pro Arg  
 395 400 405  
 Arg Gly Pro Gly Gly Gly Cys Leu Asp Val Asp Phe Ala Val Gly  
 410 415 420  
 Pro Pro Gly Cys Ser His Val Asn Ser Phe Lys Val Gly Glu Asn  
 425 430 435  
 Trp Arg Gln Glu Leu Arg Val Ile Tyr Gln Cys Phe Val Trp Cys  
 440 445 450  
 Gly Thr Pro Glu Thr Arg Lys Ser Lys Ala Lys Ser Cys Ile Cys  
 455 460 465  
 His Val Cys Gly Thr His Leu Asn Arg Leu His Ser Cys Leu Ser  
 470 475 480  
 Cys Val Phe Phe Gly Cys Phe Thr Glu Lys His Ile His Glu His  
 485 490 495  
 Ala Glu Thr Lys Gln His Asn Leu Ala Val Asp Leu Tyr Tyr Gly  
 500 505 510  
 Gly Ile Tyr Cys Phe Met Cys Lys Asp Tyr Val Tyr Asp Lys Asp  
 515 520 525  
 Ile Glu Gln Ile Ala Lys Glu Glu Gln Gly Glu Ala Leu Lys Leu  
 530 535 540  
 Gln Ala Ser Thr Ser Thr Glu Val Ser His Gln Gln Cys Ser Val  
 545 550 555  
 Pro Gly Leu Gly Glu Lys Phe Pro Thr Trp Glu Thr Thr Lys Pro  
 560 565 570  
 Glu Leu Glu Leu Leu Gly His Asn Pro Arg Arg Arg Arg Ile Thr  
 575 580 585  
 Ser Ser Phe Thr Ile Gly Leu Arg Gly Leu Ile Asn Leu Gly Asn  
 590 595 600  
 Thr Cys Phe Met Asn Cys Ile Val Gln Ala Leu Thr His Thr Pro  
 605 610 615  
 Ile Leu Arg Asp Phe Phe Leu Ser Asp Arg His Arg Cys Glu Met  
 620 625 630  
 Pro Ser Pro Glu Leu Cys Leu Val Cys Glu Met Ser Ser Leu Phe  
 635 640 645  
 Arg Glu Leu Tyr Ser Gly Asn Pro Ser Pro His Val Pro Tyr Lys  
 650 655 660  
 Leu Leu His Leu Val Trp Ile His Ala Arg His Leu Ala Gly Tyr  
 665 670 675  
 Arg Gln Gln Asp Ala His Glu Phe Leu Ile Ala Ala Leu Asp Val  
 680 685 690  
 Leu His Arg His Cys Lys Gly Asp Asp Val Gly Lys Ala Ala Asn  
 695 700 705  
 Asn Pro Asn His Cys Asn Cys Ile Ile Asp Gln Ile Phe Thr Gly  
 710 715 720  
 Gly Leu Gln Ser Asp Val Thr Cys Gln Ala Cys His Gly Val Ser  
 725 730 735  
 Thr Thr Ile Asp Pro Cys Trp Asp Ile Ser Leu Asp Leu Pro Gly

740	745	750
Ser Cys Thr Ser Phe Trp Pro Met Ser Pro Gly Arg Glu Ser Ser		
755	760	765
Val Asn Gly Glu Ser His Ile Pro-Gly Ile Thr Thr Leu Thr Asp		
770	775	780
Cys Leu Arg Arg Phe Thr Arg Pro Glu His Leu Gly Ser Ser Ala		
785	790	795
Lys Ile Lys Cys Gly Ser Cys Gln Ser Tyr Gln Glu Ser Thr Lys		
800	805	810
Gln Leu Thr Met Asn Lys Leu Pro Val Val Ala Cys Phe His Phe		
815	820	825
Lys Arg Phe Glu His Ser Ala Lys Gln Arg Arg Lys Ile Thr Thr		
830	835	840
Tyr Ile Ser Phe Pro Leu Glu Leu Asp Met Thr Pro Phe Met Ala		
845	850	855
Ser Ser Lys Glu Ser Arg Met Asn Gly Gln Leu Gln Leu Pro Thr		
860	865	870
Asn Ser Gly Asn Asn Glu Asn Lys Tyr Ser Leu Phe Ala Val Val		
875	880	885
Asn His Gln Gly Thr Leu Glu Ser Gly His Tyr Thr Ser Phe Ile		
890	895	900
Arg His His Lys Asp Gln Trp Phe Lys Cys Asp Asp Ala Val Ile		
905	910	915
Thr Lys Ala Ser Ile Lys Asp Val Leu Asp Ser Glu Gly Tyr Leu		
920	925	930
Leu Phe Tyr His Lys Gln Val Leu Glu His Glu Ser Glu Lys Val		
935	940	945
Lys Glu Met Asn Thr Gln Ala Tyr		
950		

&lt;210&gt; 22

&lt;211&gt; 2204

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID-No: 275791CB1

&lt;400&gt; 22

```

atat'ccaaat .agacctgaca :agtctgaatt ggaaaactcag attgacagaaa tgacgaagaa 60
 gagcttttagc -agctgtcttg gagataagta agagagatgc ttcaccatct ctgagtcatg 120
 .aagatgatga taagccaaact .agcagccccag ataccggatt tgcagaagat gatattcaag 180
 aaatgccgga aaatccagac actatggaaa ctgagaagcc caaaaacaatc acagagctgg 240
 atcctgcccag ttttactgag ataactaaag actgtatgatga gaataaaagaa aaaaaaaactc 300
 cagaaggatc tcagggagaa gttgttggc tccagcagta tgatatggag cgtgaaaggg 360
 aagagcaaga gcttcagcag gcactggctc agagccttca agagcaagag gcttgggaaac 420
 .agaaaagaaga tcatgacatc aaaagagcta ccgagttaaag tcttcaagag tttaacaact 480
 cctttgttggaa tgcattgggt tctgtatgggg actctggaaa tgaggatgtt tttgatatgg 540
 agtacacaga agctgaagct gaggaaactga aaagaaaatgc tgagacagga aatctgcctc 600
 attcgatccg gctcatcagt gttgtcagtc acattggtag cacttcttct tcaggtcatt 660
 acattatgttga ttttatatgac attaagaagc aagcgttggtt tacttacaat gacctggagg 720
 tatcaaaaat ccaaggaggct gccgtcgaga gtgatcgaga tggagatggc tacatcttct 780
 ttttatatgca caaggaggatc tttgtatggc tgcgtggaaa acggaaaacatc ttcgtcac 840
 ttagcacggaa agtggggaaag actaccgtc aggcctcggtg aggaacaaac tcctggggtt 900
 gcagcatgca ctgcattttt gttactgtc cccacctcactt ctttcctctg ctgaaggaga 960
 atttggaaatt ctacttgatg cgggagcaac aaacagctca gggccaaacc aaaagacaaa 1020
 aattggatgta acgttagaatg ctccatgcta ttttatggaa actttgtct cacatccgtt 1080
 gctgattatc ctctttttctt cctatgatgt gcacttcttt tgtcttagga atacatgttg 1140
 taaaatataat tctgtgtatg tttatgtataca cacacacaga cacacacaca cacacacggg 1200
 atgaatggag ccttaaagag ttaggtatgg ccaccagaat atgcctgtctc aaaattaata 1260
 gcacacgttggagaaga aatgaagggtc tcaaaagagtc cattcacgtt agaaatgtgt 1320
 gaagacatac ttatcagtgg gcttttagct tttatgttcc tttagttagtt tcactcaagt 1380
 ctgttacatctt ttgtgtttcc ttattatgtt aattcactgg aaagccagct cttcatgttta 1440
 cactaatgac agtttgttct ctttgcagaaga gagggggcatt actgtcacct gacttgaggaa 1500
 gctgttttgtt ttgtgtttttt gctctgaaat ttcatgttta tgtgtatgtt ttgtgtttta 1560
 catgcagtcc caagaaatgg attgttggtt ctttgcataa ttgtgtttttt ccacatttga 1620

```

tatttcttat atactttgtt ttctctaagg agatttttc acacagtatg ttcatcatat 1680  
 atcatcatca ttattatgtt ggttaaagata gaatctttt tcttttttgc tcattctggc 1740  
 catggagcag cattaccctt atggatttgc accaaaactt taaacaagta gaaagataat 1800  
 atttctccaa ttgggactcc ccagcaggaa tacttaggaa taaggaagaa tgcttagcatc 1860  
 tctgtctctc aaacataggg aggataagaa gagtgttctt ctggtaaagc taaaattctg 1920  
 gaccactgaa gctaaaagcc ctatttgcag tatgaaatta agtacttgag ctataggaca 1980  
 aaccttgggc atttaaccat ttactgtctg gctttgcct taaaataggg ttgcaattaa 2040  
 aatgtgattt gcttaggtaa tcccaaaaac taacaaataaa caaaggtgca taattttttt 2100  
 atctacttt tagtgctct gagttgaggg aaagtagagc ggcaacatta agtgctatgc 2160  
 tagtcactta gctgacgtaa ccagcttgg taagcagctt atga 2204

&lt;210&gt; 23

&lt;211&gt; 2036

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 1389845CB1

&lt;400&gt; 23

ccgatggggg ttaggctcca gggcttctgt cgagaccaag gatgccaaa tatctgggtg 60  
 gtgggtctg catactggg ccctggcag aacgaagggt atacagcctg gcccaccagg 120  
 ataagtccag aacccaccag gagctgagga cagacagaag gaccacggag ggggtgacgg 180  
 gctgtgtga ggatgggtc ccctggccca ggactctctt ctcttcctc tgctggctcc 240  
 agaccagagt ccaagccctt ggcagtgcctt cccttacca gcccagctt gaagacagaa 300  
 tgagagggtt tcctgtctc caggctctgc tccttcttgtt gctggagct gctgggactc 360  
 agggaaaggaa gtctgcagcc tgcggcagc cccgcattgtc cagtcggatc gttggggcc 420  
 gggatggccg ggacggagag tggcggtggc aggcgagcat ccagcatctg gggcacacg 480  
 tggcgggggg tgcgctcatc gccccccagt ggtgtctgac agcggcgcac tgcttccca 540  
 ggaggggact ggcagcttag taccggctgc gcctggggc gctgcgtctg ggctccacct 600  
 cgcggccgac gctctcggtt cccgtgcac ggtgtctgtt gccccccggac tactccgagg 660  
 acggggcccg cggcgacctt gcactgtgc agctgcgtcg cccgtgccc ctgagcgctc 720  
 gcgttcaacc ctgtgtccctt cccgtggccg ggcggccggc ggcggccggc acaccatgcc 780  
 gggtcaccgg ctggggcagc ctccggccag gagtggccctt cccagagttt cgaccgctac 840  
 aaggagaatg ggtgcgcgtc ctggacttcgc gcacctgcga cggcctctac cacgtggcg 900  
 cggacgtgcc ccaaggctgtt cgcattgtgc tgcctggag tctgtgtgcc ggctacccccc 960  
 agggccacaa ggacgcctgc caggtgtgc cccagccctt ccagcctccg gagtcccccc 1020  
 cctgtgccccca gcaccccttcc tcccttactt ccaggaccca ggacatccaa actcaggctc 1080  
 aggatcttgg cctccaaacctt agaggccca cggcagggtt ctggaaacctt gagaactgaa 1140  
 gtcctgggag ggctggact taggtcttcc ttcttctgc aggggttattt tgggggacct 1200  
 ctgacctgccc tgcagtctgg gagctgggtc ctgggtggcg tggtagctg gggcaagggt 1260  
 tggcccttgc ccaaccgtcc aggggtctac accagtgtgg ccacatatacc cccctggatt 1320  
 caggctcgcc tcaagtttctt atgctagccg gtgaggctga cctggagcca gctgctgggg 1380  
 tccctcagcc tccctgggttca tccaggccac tgccttattacc ccacatccct tctgcctcga 1440  
 gccaagatg cttaaaaaaatg ctaaaggccca ccccaccccc caccacccac ctcctgcctc 1500  
 ctctctctt tggggatcac cagctctgc tccaccaacc ctcatccagg aatctgccc 1560  
 gagtcccagg gagtcacact ccccaacttcc ttcttggctt gtatattactt ttcttggccc 1620  
 tggccaggggc tggcgcaag gcacgcagtg atgggcaaac caatttgcgc ccatctggcc 1680  
 tggtagccca tcttttctg gagaaagtca gattcacagc atgacagaga tttgacacca 1740  
 gggagatctt ccatagctgg ctttggggac acggggacca cagccatgg cggcctctaa 1800  
 gagctgagag acagccggca gggaatccgg accctcagac ccacagccgc aaggcaactgg 1860  
 attctggcag cacccttgcac gactggaa gtaagttttt ccccagcctc cagataagag 1920  
 ccccccggc caatcccttc atttcaaccc aaagagaccc taagcagaga accttagctga 1980  
 gccactccttgc accttacaaag ttgtgactta ataaatgtgt gctttaagct gctcca 2036

&lt;210&gt; 24

&lt;211&gt; 2185

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 1726609CB1

&lt;400&gt; 24

gccatgcctc ctgccccacgg ccaccagcaa gctgtcgccgc gcagtggagc agtggctgag 60

tgcagctgag cggtcttatg ggccctacat gtggggcagg tacgacattt tcttcctgcc 120  
 accctccccc cccatcgatgg ccatggagaa ccctgcctt accttcata tcttcctccat 180  
 cctggagagc gatgagttcc tggtcatcgat tgcacccatccac gaggtggccc acagttgggt 240  
 cggcaacgt gtcaccaacg ccacgtggga agagatgtgg ctgagcgagg gcctggccac 300  
 ctatccccag cggcgatca ccaccgagac ctacgggtct gccttcacct gcctggagac 360  
 tgcctcccg ctggacgccc tgcacccggat gatgaagctt ctggggagagg acagccccgt 420  
 cagcaaaactg caggctcaagg tggagccagg atgaaatccc agccacatcgat tgaacccgtt 480  
 cacctacgat aagggtactt gttcgatcgat tctatcgatcc cagctctcgat gagaccaca 540  
 ggcgtttagt gactttctcc ggcctatgt ggagaagttac aagttccatca gcgtgggtgc 600  
 ccaggacactg ctggacttcc tcctgatctt ctcccccggat ctgaaggagc agagctggta 660  
 ctggccggca gggctggaat tgcacgcgtg gtcataatggcc acaggccgc cgctggctga 720  
 gccggacactg ttcagggtat ccagctgac ccggcccgat gaggccctt tccagctgtg 780  
 gaccgcagaa cctctggacc aggacgtgc ctggcccgat gccattgaca tctccaatgt 840  
 gaggacactc cagacagcac tcttccttggat ccggctccgt gatgggtccc cgctggccca 900  
 ggaggtggat atgacgtctt ccaagtgtca ctccctcttgc ctggactcgat tgaacgtgt 960  
 gateccatc cgtggctgc agatgtgtt ccgaaacgac tactatcttgc acctccacag 1020  
 ggtggccgc ttcttggaga gccagatgtc acggatgtac accatccgc tgcgtggagg 1080  
 cctctgcacc ggtgcctca agtcctcgat gctggaggtt ttcttccatca cgcaggcccg 1140  
 gtcgcacccc aacctgcgcgca gagccatcca gcaatccgt tcccaaggcc tgggtccat 1200  
 cacagagccc gcctcagagc ccagcacggat gctggggcaag gctgaaggacg acacagactc 1260  
 ggacgcacag gcctgtgc ttggggacga ggccccccagc agtgcacatct ctctcaggga 1320  
 cgtcaatgtc ttcgtcttagc cctgtggcg ggctgacccctt cgaccccttca gacaccacaa 1380  
 ttgtgccttc tttggggccat gcctgcatcg actgcgtctc ggctctggcc atgagctctg 1440  
 cccaggccca caaggccctc ccctggctc tccaggcag ggagaatggg gagaggggacc 1500  
 tccttgtgtc tggcagagac ctgtggaccc tggctccatca ctcccagctc tcttcactg 1560  
 caggccctgg gggcagcccg cacacaccat gcttcctgtc tcaacactga cagctgtgcc 1620  
 tagccccggat tgccagcacc tgccaggatgc cgccccgggg caagggcccc agcagcccta 1680  
 tggtgaccgc cacactgtgc cttaatgttgc gccccggggcc caggctgtgc tttccctgc 1740  
 gcacgccttc ttgcagggtat ctgcacccatccccc gacccctgcac cccggcccta 1800  
 gggttggcag cttcagttgg cccctggcagg aggaacaagg acacagacat tccctcagt 1860  
 tggggggcag gggacacagg gagaggatgg ttgtccctgg ggagggccct ctggcccccag 1920  
 gcaacccatgg cccctcagaa caggatgc caggacccag ggagatgtg gggacaggac 1980  
 acgtgtcttc ttgttagttc ctgggtggggaggcagagg gcaaaagcaat accccaggga 2040  
 aagtggggagg tgggtgtgtt gctctccatccatccat gctggggagag gcggcccgag 2100  
 cctggggctt ccaggctggg actgctgtatca tggggatgttca cggtgatggt cccattaaac 2160  
 ttccactctg .aaaaaaa aaaaa 2185

&lt;210&gt; 25

&lt;211&gt; 3486

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 4503848CB1

&lt;400&gt; 25

ctgtttttaaa aaaagggggat gggaaagatca ctgatttat tttataaaatggaa gaattttat 60  
 agctccaaca cctgacttta tttatctata tggtttattt acacaaacaa ttcaatgtttt 120  
 gaattataca aatttcatta aaactatgtt attatgtca aaaaatggaa aatacagggg 180  
 cacttagttt tttgtcatatg ttttccatcg agatgtttt tttttttttt taaaatggtt 240  
 tttttttaattt taatattttaa ttttataatgtt cacatttata ttttgcattttt ggaccaacag 300  
 aaatccctaa ttcttattttt ttttctgtt tttttttttt tttttttttt tttttttttt 360  
 ctcagaatgtt ctttgcattttt tttttttttt tttttttttt tttttttttt tttttttttt 420  
 gttggctggca gaaatgtttt aatgtggat ttttccatcg agccaaatcaa aacaagaact 480  
 gaggagctca ttgtgcagac caaacatgtt tacgtatgtt tttttttttt tttttttttt 540  
 gaagtaactt acggaaactt ttttgcattttt tttttttttt tttttttttt tttttttttt 600  
 gaaaggatca ttttgcattttt tttttttttt tttttttttt tttttttttt tttttttttt 660  
 agtacagaatg ctttgcattttt tttttttttt tttttttttt tttttttttt tttttttttt 720  
 atatttgatca gaaatgtttt ttttgcattttt tttttttttt tttttttttt tttttttttt 780  
 gttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 840  
 gttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 900  
 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 960  
 gttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 1020  
 acctttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 1080  
 agaaggatgg aatggctttt tttttttttt tttttttttt tttttttttt tttttttttt 1140  
 cttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 1200

ttcgtcccttg	aaatgaacac	tgcaagagc	acaaggccg	taacagccct	tcttagatgat	1260
ttaaggccaga	agttaaaacc	cttgggtgaa	gcagaacag	agtttattt	gaatttgaag	1320
aaaaaggaaat	gcaaagacag	gggtttgaa	tatgtggga	aaatcaatgc	ctggatcta	1380
tattactaca	tgactcgac	agaggaactc	aagtattcca	tagaccaaga	gttcctcaag	1440
gaatactcc	caattggat	ggtcaactgaa	ggcttgcgt	acacccatcca	ggagttgtt	1500
ggactttcat	ttgaacaaat	gacagatgt	catgttgg	aaagaggtgt	tacacttat	1560
actgtgaagg	ataaaagctac	aggagaagta	ttggacagat	ttatggg	cctctatcca	1620
agggaaagaa	aatacaatca	tgcggccctgc	ttcggtctcc	agctggctg	ccttctgcct	1680
gatggaaagcc	ggatgtatggc	agtggctgcc	ctcggtgt	acttctcaca	gccagttggca	1740
ggtcgtccct	ctctcctgag	acacgacgag	gtgaggactt	actttcatga	gtttggtcac	1800
gtgtatgcac	agatttgtgc	acagactgt	tttgcacgt	ttagcggaa	aatatgtggaa	1860
actgacttt	tagaggtgcc	atcgccaaatg	cttggaaaatt	gggtgtggga	cgtcgattcc	1920
ctccgaagat	tgtcaaaaa	ttataaaagat	ggaagcccta	ttgcagacga	tctgttgc	1980
aaacttgtt	cttcttaggct	ggtaacacaca	ggtcttgcgt	coctggccca	gattgtttt	2040
agcaaaagg	atcagtctct	tcataccaaac	acatcgctgg	atgtcgaag	tgaatatgcc	2100
aaataactgtct	cagaaatatt	aggagttgca	gctactccag	gcacaaatat	gccagctacc	2160
tttggacatt	tggcagggggg	atacgatggc	caatattatg	gatatcttt	gagtgaagta	2220
ttttccatgg	atatgtttt	cagctgtttt	aaaaaagaag	ggataatgaa	tccggagggtt	2280
ggaatgaaat	acagaaaacct	aatcctgaaa	cttggggat	ctctggacgg	catggacatg	2340
ctccacaatt	tcttggaaacg	tgagccaaac	caaaaacgc	tcctaattgag	tagaggcctg	2400
catgctccgt	gaactgggg	tctttgttag	ccgtccatgt	ctggaggaca	agtcgacatc	2460
accatgtgtt	actggcctgg	aaactgaagg	gagtttgc	atgaaaatt	tagatttcta	2520
tttgacatct	tttggtttct	aattttaaaa	attataaaaga	tgtaaatgga	attataaaata	2580
ctgtgaccta	agaaaaagacc	cactagaaag	taattgtact	ataaaaattt	ataaaaactgg	2640
atttgattt	tttttatgaa	agtttcatat	gaatgtact	tgatttttta	ctattataat	2700
ctagataata	tgtatataaga	gggtcaagaa	tttttaaatt	gaatcatata	tatgatataaa	2760
tttgcattt	cttgcattt	gaagttttgt	acttgggtt	tctggactga	taaatgaatc	2820
atcacattt	tctgttaat	atttttttgg	agctctgt	caactttgt	cctttgtct	2880
ccaggaaggt	gtgacctctc	ctttgcctgc	atacctcaag	gccaggggaa	tatgcctcag	2940
tgatgcattt	atctttgtat	atcaggccgc	atgatttcca	actttctgcc	acacttaaat	3000
tacgttcc	catttcagtt	ttgtctttt	tgtctaaagt	tcaagtcaaaag	agtatcaaaa	3060
aattatgtt	cagcttagact	ggtgtaatgt	ataagtttt	gtatcttgc	ttagaggatt	3120
tcgttagtt	tattagaggc	tcattttccac	ctcagcatac	aagatcgta	gtcttttggc	3180
atgtgtgc	attagaaatac	taaaggcaagt	ccaagcacat	tttcttc	tcacgtttct	3240
aataaagtgtt	aggactttg	cctcttttac	ttacacgtc	ccccaaaatgt	tcaggttagac	3300
atgtcacaaa	tggctctgt	gagagccatg	ggaagagaga	ggaggtggat	gtggacata	3360
aagggttca	aaactccaga	agaggagtgg	gtttttggata	gaagacatgtt	aggacagctg	3420
ctccaaagcc	ttatgtgtat	gatgaaactt	aaccacgggg	aagagactct	tcaagttagct	3480
gttctg						3486

<210> 26 .

<211> 2847

<212> DNA

<213> Homo sapiens

<220>

<221> misc\_feature

<223> Incyte ID No: 5544089CB1

<400> 26

caatgacgct	tggacgagga	tttatttcta	caagctaatt	gaatccagga	gcagctttaa	60
ttataaacac	taacggaaga	aaaaaggagt	atttccaagg	gctcaaatgg	aagctgtact	120
cagtcgggt	gaggcagggg	gaggtaaagt	ttctcacact	caagtctgt	tcatagttt	180
ctgtccctt	ccaaacaaaa	gctaataacg	ccatacgcat	ccacacactc	cctcctggat	240
gaacacta	ctcgcccc	ctgtcacccc	aaggccagtt	atcaaaaact	gttccttctc	300
tgccctcaaa	gactgaaggg	cgaggccctg	tttgcctt	gtcaggat	ctgattgtct	360
ttaaagtgt	cttacaagat	tccgtcgat	tttgcctt	ctgtcttgtc	atcaggacta	420
agtggtgagg	catcaaaaagg	tagaaagatg	gaacttattc	agccaaaggaa	gccaacttca	480
cagtacattt	ctctttgtca	tgaattgcat	actttgtcc	aagtcatgtg	gtctggaaag	540
tggcggttgg	tctcaccatt	tgctatgcta	cactcagtgt	ggagactcat	tcctgccttt	600
cgtggttacg	cccaacaaga	cgctcaggaa	tttctttgt	aacttttaga	taaaatacaa	660
cgtgaattag	agacaactgg	taccatgtt	ccagcttta	tccccacttc	tcaaaggaaa	720
ctcatcaaa	aagtgtgaa	tgttgtaaat	aacattttc	atggacaact	tcttagtgcag	780
gttacatgtc	ttgcatgtca	caacaaatca	aataccatag	aaccttttgc	ggacttgtca	840
ttggagttt	cagaaaggta	tcaatgcagt	ggaaaaagata	ttgtttccca	gccatgtctg	900
gttactgaaa	tgttgccaa	atttacagaa	actgaagctt	tagaaggaaa	aatctacgt	960
tgtgaccagt	gtaaactcaaa	ggctgagaagg	ttttcccca	aaccaggatgt	actcacagaa	1020

ggccagaaaac	aacttatgat	atgccaccta	cctcagggttc	tcagactgca	cctcaaacga	1080
ttcagggtgt	caggacgtaa	taaccgagag	aagattggtg	ttcatgttgg	cttgaggaa	1140
atcttaaaca	tggagcccta	ttgctgcagg	gagaccctga	aatccctcag	accagaatgc	1200
tttatctatg	acttgtccgc	ggtggtagt	caccatggga	aaggatttgg	ctcagggcac	1260
tacactgcct	actgtataa	ttctgaagga	gggttctggg	tacactgcaa	tgatctccaa	1320
ctaaggatgt	gcaactatgg	taagaatgc	aaggctcaag	cttatactt	gttttatacc	1380
caacgagtt	ctgagaatgg	acattctaaa	ctttgcctc	cagagctcct	gttggggagc	1440
caacatccca	atgaagacgc	tgataacctcg	tctaattgaaa	tccttagctg	atccaaagac	1500
aatggggttt	tcttcctgtg	atttataatat	atactttttt	aaagactgat	gtaccatttt	1560
aaacttcatt	ttttcttgtg	aatcagtgt	tactacattt	atacatttt	tatccaacaa	1620
ttttttttt	tacaaaagtat	aaatgtat	atcaactgaa	ggttaactact	ttttcatat	1680
ttggagttt	aaacttttg	tgtttacctc	agactgtatgt	taccttttt	atattttat	1740
gtcttaattt	gctcgatgt	tgaacttgt	caatctctaa	ccaacaaagt	tcaagtggca	1800
tcattttata	tacatgtatc	ttttccagg	attttctata	caaatttta	atagatggaa	1860
aatttagactc	tactttggtc	actaataatgc	tttcattttgt	atattgaagt	taccttggcc	1920
cttggagtt	ttgaagtgtac	atgtcaaggt	atccacctaaa	tattcttcag	tcacactcac	1980
tggtattttct	gaggctttgt	gtgttaacag	gccttgcataat	tgacattatt	ttgggttaatg	2040
taaccccaaa	attgtctttag	taattgtct	ttggcatag	caaactataa	atgaaaatgg	2100
cagctttaca	aatagttatata	ttaagtgtac	tctggaaacta	tggacatgaa	aaaaaatgtat	2160
gctgggattt	atgatttttt	tctggcagca	acaagggttt	tccagaagtc	taataattaa	2220
gcagtcataaa	aaagtctgaa	ttagttaaac	cagtgtat	tgttatttcaa	atagtttacc	2280
ttgggtatga	gttcatttttta	taatgtctga	tgacattaga	tctctttaaa	ctttatgtat	2340
tttttttagt	tcaaaggaaat	agagtcttga	agagaaaaaa	ttatagggca	gaaaagataa	2400
gtgttcaaaa	ttggcaactg	gactattatt	atgtctagca	tctcattct	aataactaaa	2460
gcttgattt	ctcttgcgt	gattatgtga	ctacttaggt	ggagccctt	aaaacactgg	2520
ccctgagcat	aaaaaaaaaa	aaaaaaaaact	aaaagctatc	tatcttaact	tgcaaaaaaaa	2580
aaattecgg	gggggttcacc	cttttcttc	tctctttaaaat	ctcacggggt	ttttttaaaag	2640
-cccttgcgt	ggggacttta	tcttttttgg	ggggggtaga	atcacctaat	ctctgtagac	2700
eagctatgtt	tctaagctct	gttaaccacg	ggggatctg	gtacccttt	tttaaaagg	2760
•gggtttttttt	ccccgggttgaag	tcttagtggaa	aagtagtccc	ctggagaatg	cggtccaccc	2820
•ctggggggccca	tctgttaggt	aaaactt				2847

4<210>27

<211> 890

...<212>>DNA

<213> Homo sapiens

•  
•  
**<220>**  
•  
•

<221> miscfeature  
<222> InputID No

<223> Incyte ID No: 7474081CB1

•<400> 27

```

>gaggcacaaga atccggcacg aggcaattac tccctgactt aaaaaaaaaaa 60
>ccatggaaata ttgttttttat ttgggtgtcc tcgctggac atttttttttt gctgactcat 120
>+ctgtttcagaa agaagaccct gctccctatt tggtgtacct caagtctcac ttcaaccctt 180
>+gtgtgggcgt cctcatcaaa cccagctggg tgctggcccc agctcaactgc tatttaccaa 240
>+atctgaaaagt gatgctggga aatttcaaga gcaagactcg agacggtaact gaacagacaa 300
>ttaaccccat tcagatcgtc cgctacttgg actacagtca tagcccccca caggatgacc 360
>tcatgctcat caagctggct aaggcctgcca tgctcaatcc caaagtccag ccccttaccc 420
>tcgcccaccac caatgtcagg ccaggcaactg tctgtctact ctcaaggttt gactggagcc 480
>aagaaaaacag tgcccgacac cctgacttgc ggcagaaacct ggaggcccccc gtgtatgtctg 540
>atcgagaatg cccaaaaaaaaca gaacaaggaa aaagccacag gaattcccta tgggtgaaat 600
>ttgtgaaaagt attcagccga atttttttggg aggtggccgt tgctactgtc atctgcaaaag 660
>acaagctcca gggaaatcgag gtggggcact tcatggggagg ggacgtccgc atctacacca 720
>atgttacaa atatgtatcc tggatttgaga acactgtcaa ggacaagtga gaccctactt 780
>cctccctctgc attccactgg ctctgcattg gactatacaa gcagataatt ttccctcttat 840
>tcaaaaataaa atctccaaaat gaaaatttgg gaatgttagca aaaaaaaaaaa 890

```

<210> 28

<211> 1577

<212> DNA

<213> Homo sapiens

<220>

<221> misc\_

<223> Incy

&lt;400&gt; 28

atgcagccca cggggccgcga gggttcccg gcgcgtcagcc ggccgttatct gggcgctctg 60  
 ctgcctctgc tactgtctgt gtcgtcgccg cagccccgtaa cccgcgcgga gaccacgcgc 120  
 ggcgcggccca gagccctctc cacgtggcc tccccccagcc tcttaccac gcccgggtgtc 180  
 cccagcgcggcc tcaactacccc aggcctcaact aegccaggca cccccaaaac cctggacacctt 240  
 cgggtcgccg cgcaggccct gatgcggagt ttcccactcg tggacgcca caatgacactg 300  
 ccccaaggcc ttagacagccg ttacaagaat gtgcgtcagg atgttaacct gcgaaatttc 360  
 agccatggtc agaccaggctt ggacaggctt agagacggcc tcgtgggtgc ccagttctgg 420  
 tcagcctccg ttcatgcctca gtcggccaggac cagaactccg tgccgcctcg cctggagccg 480  
 attgacctca ttcaccgcata gttgcctcc tactctgaac tcgagcttgt gacctcagct 540  
 gaaggctctga acagctctca aaagctggcc tgcctcattt ggcgtggaggg tggtcactca 600  
 ctggacagca gcctctctgt gtcgcgtcgt ttctatgtgc tgggggtgcg ctacactgaca 660  
 cttaccttca ctcgcgtac accatggca gagagttcca ccaagtcag acaccacatg 720  
 tacaccaacg tcagccggatt gacaagcttt ggtgagaaag tagtagagga gttgaaccgc 780  
 ctggcatgtga tgatagattt gtcctatgcata tggacacacct tgataagaag ggtcctggaa 840  
 gtgttcagg ctctgtgtat cttctcccaact tcagtcgtcc gacgtgtgt tgacaatttg 900  
 ttgaatgttc cctgcgtat cctgcgttccctt ctaagaagaactt acgggtgcatt cgtgtgggt 960  
 acactgttca tgggggtgtc gcagtcgttccact tcgttgcata acgtgttccac tggccggat 1020  
 cacttgcacc acatcaggggc agtcatttgcata tggagttca tcgggatttgg tggaaatttt 1080  
 gacggactg gccgggttccc tcagggctg gaggatgtgt ccacataccca agtctgtata 1140  
 gaggagttgc tgagtcgttag ctggagcggag gaagagctt aaggtgttcc tcgtggaaac 1200  
 ctgtcgccgg tcttcagacata agtggaaaag gtgagagagg agacggggc gcagagcccc 1260  
 gtggagggtc agtttccat tgggcaactt acacatctt gccactccca cctcgtgcct 1320  
 cagaatggac accaggcttca tcatctggat gtcggccaggac agccaaccaa tgggtcccc 1380  
 tggaggttcc taaaatgcctc cccatcacctt gtccggcc tttgtgtc tgccaccatc 1440  
 ccaaccttca cccatgggttctgatgcgttcc acgtgttcccg cgcagaggc actgtggcaa 1500  
 agcctcacaa agccccctctt cctagttcat tccacaaggat atgctgagaa taaacatgtt 1560  
 acacatggaa aaaaaaaaaaaaaaa 1577

&lt;210&gt; 29

&lt;211&gt; 1958

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 2256251CB1

&lt;400&gt; 29

aagccgtcga gtcggcatt cattgttaacg ggcgcattgtg ctggaaagggt cgtgtggttt 60  
 ctgcgtcgat ctctcggtt agtggggctg gtcgggggtgt gtcgcaggcc tgcgtccccc 120  
 accaccactg tagtcgttot gtaccttggg agatgtctg aggccatgaa acaacactggc 180  
 ctcctcgaa ctttctccccc acagcgtccc caccgtggcc ctggaaaccag ctgggggttt 240  
 tggcggtgtgg gagagccgggt gccccagccc acaccgcgtg cctatctata aacatctcg 300  
 tctgttacata tcccgatcttc cttccatcc agcccgatgg gcaacttccca tcaccagcac 360  
 aattatccat ctcaggcaga cccaggatgtg gccagggtgtt ctccagatgtt cttcccttta 420  
 atttcttctt aacttactgt ggtgaagttt acagaagata aagttacatt atcaagcgtc 480  
 caattcagag gccccggagca ctttggacag tgctgtctgc ccccccggcccc cggaaatttt 540  
 tccagcttca aatatctcca ccacccttga agaaaaaccg gggcccaacta ggcagccacc 600  
 cctagcaccc cggggccttot cgggggctcc accgttctga gcccctttt agccgcctag 660  
 gggccctctg cggccctttcc accggccctcg gtagccctgg ttagtttgg gaggatgtgt 720  
 cttagaaaacg acgacatgtt gcccaggccg gtcactgtt cctggagacgt catttctgt 780  
 gcaccacacca tgcttctggg gagatgtgg cagacggagag agtgaagag caaatggccc 840  
 aagaaggcag ggaggtgtgg tcagggaggatcttgc gtcgttccacttccatggatgttcc 900  
 ggtatccccac caggcaccatc ttcccttcc gactttagggt gtggccggcc gcagggtttcg 960  
 gatgcaggccg gccggatctgtt ggggggtcac gtcggccccc ccggccgtat gccatggcag 1020  
 gccagccctcc gccttgcggag ggtgcacgtg tgccggccgtt cactgtctcg ccccccgtgg 1080  
 gtgcgttccacag ctggccacttccctccggg tccctgtacttccatccatccatccatccatcc 1140  
 ctgggggaatc tggagatcac tctgttcccttccacttccatccatccatccatccatccatcc 1200  
 cacttcaggcc ccttcaggacca gccggggacc agccggggaca tccggccctgttggatgttcc 1260  
 gtcccccgttca cccttcttccatccatccatccatccatccatccatccatccatccatccatcc 1320  
 ttctgtccctg gatccgggtt ctgggttgcacc ggttggggctt atacggggaa gggagggctt 1380  
 ctggccaccatccatccatccatccatccatccatccatccatccatccatccatccatccatcc 1440  
 cgccggggact atccggccatccatccatccatccatccatccatccatccatccatccatccatcc 1500  
 ggccccggggatc tggcccttccatccatccatccatccatccatccatccatccatccatccatcc 1560  
 gccttgggtgc aggtgttccatccatccatccatccatccatccatccatccatccatccatccatcc 1620  
 ggatgttccatccatccatccatccatccatccatccatccatccatccatccatccatccatccatcc 1680

```

gggggctcg agtctgggtt ccccaggctc ccccttcctgg ctggcttatt cctccccggc 1740
ctttcccttc tgctagtctc ctgtgtccctg ctggccaagt gcctgtgcg cccatctgcg 1800
gatggtactc ccttcccccgc ccctgactga tggcaggaat ccaagtgcgat ttcttaata 1860
agttactatt tattccgctc cgccccctcc ctctcccttg agaagctgag tcttctgcgat 1920
cagattatttgc caacatttaa cctgaattta acgacacc 1958

```

<210> 30  
<211> 3106  
<212> DNA  
<213> *Homo sapien*

<220>  
<221> misc\_feature  
<223> Incyte ID No: 7160544CB1

atattcctga gagaccgc aataccataa gaattactaa aaaaaaa

3106

<210> 31  
<211> 3567

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7477386CB1

&lt;400&gt; 31

atggctccac tccgcgcgct gctgtctac ctgctgcctt tgcaactgtgc gctctgcgcc 60  
 gcccggggca gcccggacccc agagctgcac ctctctggaa agctcagtga ctatggtgtg 120  
 acagtgcctt gcagcacaca ctttcgggaa cgcttcctct cccacgtgtt gtctggccca 180  
 gcagcagccct ctgcaggggat catggtagtg gacacgcac ccacactacc acgacactcc 240  
 agtcaacctcc gggtgtctcg cagccctctg caccaggag ggaccctgtg gcctggcagg 300  
 gtggggcgcc actccctcta cttcaatgtc actgttttgc ggaaggaaact gcacttgcgc 360  
 ctgcggccca atcggagggtt ggttagtgcac ggatcctcag tggagtggca ggaggatttt 420  
 cgggagctgt tccggcagcc cttacggcag gagtgtgtt acactggagg tgtcaactgga 480  
 atgcctgggg cagctgttgc catcagcaac tgtgacggat tggcgggcct catccgcaca 540  
 gacagcaccg acttcttcat tgagcctctg gagcggggcc agcaggagaa ggaggccagc 600  
 gggaggacac atgtgggtgtt cccggggggc gcgtccagc aggagtgggc agaacctgac 660  
 ggggacactgc acaaatagaac ctttggctgt ggagacatcc ccaacactgtt gggctgggt 720  
 ggggaccacg tgggcacac agagcggaa cggcggcatg ccaagccagg cagctacagc 780  
 atcgaggtgc tgctgggtgtt ggacgactcg gtgggtcgct tccatggcaa ggagcatgtg 840  
 cagaactatg tcctcaccct catgaatatac gtggtagatg agatttacca cgatgagtcc 900  
 ctgggggttc atataaaatat tgccctcgic cgcttgatca tggttggcta ccgacagcag 960  
 tccctgagcc tgatcgagcc cgggaaaccc tcacgcagcc tggagcagggt gtgtcgctgg 1020  
 gcacactccc agcagcggca ggacccccc caccgtgac accatgacca cgttgtgttc 1080  
 ctcacccggc agactttgg cccctcagggt ggttatgcac cctgtactgg catgtgtcac 1140  
 cccctgagga gctgtccctt caacatgtgatggctt cctcagcctt cgtgatagct 1200  
 catgagaccc gccacgtgtc cggcatggag catgacggc acgggaatgg ctgtgcagat 1260  
 gagaccaccc tgggcagcgt catggccccc ctgggtgcagg ctgccttcca ccgcctccat 1320  
 tggtcccgct gcaagcaagct ggagctcagc cgttacccctc cgttctacga ctgcctccctc 1380  
 gatgacccct ttgatcctgc ctggccccag ccccccagagc tggctggat caactactca 1440  
 atggatgagc agtgcgcgtt tgacttggc agtggctacc agacactgtt ggcattcagg 1500  
 accttggc cctgcagaacca gctgtgggtc accatccctg acaacccgtt ctctgcaag 1560  
 accaagaagg gggcccccgtt ggtatggact ggtgtgcac cggcgaagggt gtgttcaaaa 1620  
 ggtcaactgca tctggaaatgc gccggagcag acatatggcc aggatggagg ctggagctcc 1680  
 tggaccaagt ttgggtcatg ttgcgggtca tggggggcg ggggtgcgatc ccgcagccgg 1740  
 agctgcaaca acccctcgcc agectatggc gggccctgt gcttagggcc catgttcag 1800  
 taccaggtct gcaacacgcgaa ggagtgcctt gggacctacg aggacttccg ggcccagcag 1860  
 tggcacaagg gcaactctca ctatgtgcac cagaatgcca agcacagct ggtgccttac 1920  
 gagcctgacg atgacgccccaa gaagtgttag ctgatctgcg agtgcggccga cacgggggac 1980  
 gtgggtttca tgaaccagggt gtttcacgtt gggacacgtt gcaagctaccg ggacccatac 2040  
 agcgtctgtc cggcgtggcgtt gttgtgttgcgtt gtcggctgtt acaaggagggt ggggtccatg 2100  
 aaggccggatg acaagtgtgg agtctgcggg ggtgacaact cccactgcag gactgtgaag 2160  
 gggacgctgg gcaaggccctt caaggcggca ggtgctctca agctgggtca gatcccagca 2220  
 ggtgcaggc acatcccgat tgaggcactg gagaagtccc cccaccgcatt tgggtgtaaag 2280  
 aaccaggctca cccggcaggctt catccctcaac cccaaaggccaa aggaagccac aagccggacc 2340  
 ttacccggca tggggcttggaa gttggggat ggggtggggg atgccaaggaa aaggctcaag 2400  
 accaggccggc cccctgccttga agccatgtcc atctggctc tcccccaac tgagggtggc 2460  
 ccccgccaggca gcctggcccttca caagttacgtc atccatgggg acctgtctcc cttatcggtt 2520  
 agcaacaatg tgctccctggaa ggagatggac accatgtggat gggcgcttcaa gagctggggcc 2580  
 ccctgcaggca agggctgtgg aggaggatcc cagttccacca aatacggctt ccggcgcaga 2640  
 cgagaccacc acatgggtca ggcacacccgt tggaccacca agaagaggcc caagccatc 2700  
 cggccggcgtt gcaaccaggca cccgtgtctt cagcgtgtt ggggtgacggaa ggagtgggg 2760  
 gcctgcaggcc ggagctgtggt gaaagctgggg gtgcagacac gggggataca gtgcctgtg 2820  
 ccccttcata atggaaacccttca caaggctcatg cccggccaaat cctgcgcggg ggacccgctt 2880  
 gaggccggcaggcc gggccctgtt cccgtgtcc tggccggccccc agtggaggctt gggggcttgg 2940  
 tccctgtgtt ctggccacccctt gggaggggc atccaggcgc ggcagggtt gtgcaggacc 3000  
 aacggccaaaca gcctcgggca ttgcggggg gataggccag acactgttca ggtctgcagc 3060  
 ctgcccgcctt gtggagcggtt gcccctgcacg gggagacagggt ctgtcttcgtt ccagatggaa 3120  
 gtgctcgatc gttactgttcc cattcccgcc taccaccggc tctgtgtgtt gtctgcattc 3180  
 aagaaggccctt cggggcccccacca ccctggccca gaccctggcc caaccteact gcccccttc 3240  
 tccactccctg gaagccctt accaggaccc caggaccctt cagatgtgc agagccctt 3300

ggaaagccaa cgggatcaga ggaccatcg catggccgag ccacacagct cccaggagct 3360  
 ctggatacaa gtc(ccc)agg gacccagcat ccctttgccc ctgagacacc aatccctgga 3420  
 gcatctgga gcatctcccc taccacccc ggggggctgc cttggggctg gactcagaca 3480  
 cctacgcccag tccttgagga caaaggccaa cctggagaag acctgagaca tcccgccacc 3540  
 agcctccctg ctgcctcccc ggtgaca 3567

<210> 32  
 <211> 2930  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 7473089CB1

<400> 32  
 cacgcagacc ggggcagcgg ccgagagccc gccccagccc cttcccacag cgccgcgttg 60  
 cgctgcccgg cggcatgctt ctgctggca tcctaaccct ggcttcgccc gggcaaccg 120  
 ctggaggctc tgagccagag cgggaggtag tcttccat ccgactggac cggacattta 180  
 acggccgccc ctactactgg cgggttcccgg aggactccgg ggtcagggta ctattttc 240  
 agatcacacgc attcaggag gactttacc tacacctgac gccggatgtc cagtcttgg 300  
 ctcccgcctt ctccacttag catctggcg tccccctcca ggggcttacc gggggcttt 360  
 cagacctgcg acgtcttc tattctgggg acgtgaacgc cgagccggac tcgttcgtg 420  
 ctgtgagcct gtggggggg ctccgcggag ctttggcta cggaggccgc gatgtgtca 480  
 ttagccgcgt gccaatgtc agcgcggccg cggcgcagcg caacagccag ggcgcacacc 540  
 ttctccagcg cgggggtgtt cgggggggc ctccggaga ccccacctc cgctgcgggg 600  
 tggcctcggg ctggAACCC gccatccatc gggccctgga cccttacaag ccgcggccgg 660  
 cgggcttcggg ggagagtgcgt agccggcgca ggtctggcg cgccaaagcgt ttctgttca 720  
 tcccgccgtt ctggagagacg ctgttggcg cggacgatc aatgttcaag ttccacggcg 780  
 cggacttggaa acattatctg ctgacgtgc tgcaacaggc ggcgcgatc taccggccatc 840  
 ccagcatctt caaccccatc aacatgtt tggtcaagggt gctgttctt agagatgtg 900  
 actcggggcc caaggteacc ggcataatcgcc cccctgacgtc ggcacacttc tggtcctggc 960  
 agaagaaggctt gaacaaaagggt agtgacaago accccgagta ctgggacactt gccatcttct 1020  
 tcaaccaggca atggatgtgtt ggagccacca cctgtgacac cctggcatg gctgtatgtgg 1080  
 gtaceatgtgtt gaccccaag agaaggtgtc tggatgttcaatgaa ggacgttggg cttccatcag 1140  
 ccttcaaccatc tgcccaagag ctggccacgc tttcaatc gccccatgac aatgtgaaag 1200  
 ttttgttgggg aagctccgg ccaaccatc gatgttcccg accctcatcc 1260  
 gagatcgaccgtt gccaaccctt ggtcaatgc tggatgtctc catcatcacc gacttctgg 1320  
 gagcggggca atggatgtt gactgc ctctgttgc aaccccgacaa gccatctcc ctggccgggg 1380  
 atctggggggc gecagctac accctgagcc agcagtgcga gctggctttt gggtgggtt 1440  
 ccaageccctgtt tccatcatgtt cagtaatgc ccaagctgtg gtgcaccggg aaggccaagg 1500  
 gagatgtgtt gtgccagacc cggccacttcc cctggggccgt tggcaccagc tggccggagg 1560  
 gcaatgtctg cctcaaaagggt gctgtgttggg agagacacaa cctcaacaag cacagggtgg 1620  
 atgggttctgtt ggcataatgg gatccctatg gcccctgtc ggcacatgt ggtggggcg 1680  
 tgcacgttggc caggaggcag tgaccaacc cccaccctgc caacggggc aagtactgc 1740  
 agggaggatgg ggttgaatatac cgtatgtca atctggagcc ctggccccagc tcagecttccg 1800  
 gaaaagagctt ccggggaggag cagtgttggg ctttcaacgg tacaaccac agcacaacc 1860  
 ggctcaactt cggccgtggca tgggttgcacca agtactccgg cgtgttccc cgggacaagt 1920  
 gcaagctcat ctggcgagcc aatggcactg gctacttcta tggatgttggca cccaaagggtgg 1980  
 tggatgttggc caccgtgttc tcttctgtact ccacccctgtt ctgtgttccaa ggcaagtgc 2040  
 tcaaggctgg ctgttgcgttggg aacctgggtt ccaagaagag attcgacaag tgggggtgtt 2100  
 gtggggaga caataagacg tgcaccaagg tggactggact ctccaccaag cccatgcatt 2160  
 gctacaattt ctgttggccg atcccccggc ggcctcaag catgcacatc cggccggcg 2220  
 gtttacaaaagg gctgtatggg gatgaccaact acctggctt gaaaaacagc caaggcaagt 2280  
 acctgttca cggccatttc tgggttgcgg cgttggagcg ggacctgggtt gtggaggcc 2340  
 gtctgttgcg gtacagccgg acggggacacag cgttggagag cctgcaggct tcccgccca 2400  
 tcttggagcc gtttgcgttggg gaggttccctt cctgtggggaa gatgacaccg ccccggttcc 2460  
 gctacttctt ctatctgccc aaagacccctc gggaggacaa gtccttcata ccccccgcacc 2520  
 cccggggagg accctctgtc ttgcaccaaca ggttccctcag cctctccaa cagggtggagc 2580  
 agccggacga caggccccctt gcaacgttggg tggatgttggcag ctggggggccg tggatgttggca 2640  
 gctgttggcag tggccctgtca aacggggccgg tggactggcg gggatgttggcc gggcagccca 2700  
 cggccctgc ctgttgcgttgc gccatccggc cctgtggagac acaagctgc gggggggccct 2760  
 gccccacctg ggatgttgcg gctgttgcac cctgttccaa gagatgttggcc gggggatttc 2820  
 agaggcgctc actcaatgtt gttttttttt gggccggct gttttttttt gggccggcc 2880  
 atttgcaccgc caagcccccggc gagatgttgc tttttttttt gggccggcc 2930

<210> 33

<211> 4230  
<212> DNA  
<213> Homo sapiens

<220>  
<221> misc\_feature  
<223> Incyte ID No: 7604035CB1

&lt;400&gt; 33

agcgagggtt cctggagaga ggcgcctggc gcagaagggt taacgggcca cggggggctc 60  
 gcagagcagg aggggtctc cgacgggtgt gtcccccact gcactcctga acttggagga 120  
 cagggtcgcc gcgagggaacg cagagacac cttccacgac cagatgcctg cgtagtttt 180  
 gtgaccagtc cgctcctgcc tccccctggg gcagtagagg gggagcgtg gagaactgga 240  
 ctggcaggcc ctggctgtat ctgtctgtc ttctgtccct ccctcagtc tgcttggatc 300  
 aggaggttgtt gtcggacac taccttcaga cacctacaga ggagggccag ggccccgaag 360  
 gtgtctgggg accttgggtc cagtggcct ctgtctccca gcccgtcggg gtgggggtgc 420  
 agcgaggag cccgacatgt cagctcccta cagtgcagct ccacccgagt ctggccctcc 480  
 ctcccccggc cccaagacat ccagaagccc tcctcccccgg gggccagggt cccagacccc 540  
 agacttctcc agaaaccctc cccttgataa ggacacagtc tcggggaaagg gttggccac 600  
 ttcgagggtcc cgcttccac ctaggagag aggagaccca ggagattcga gcggccagga 660  
 ggtcccggtc tcgagacccc atcaagccag gaatgttcgg ttatgggaga gtggcccttg 720  
 cattggccact gcacccggaa cgcaggacc ctccggagccc accccatct gagctgtccc 780  
 tgatcttc tagagggaaa gggccttccatc ctggggatc tccaaagcga gagccatttc 840  
 ccccaaacgg cagcccccaactgagctcc tcctccacaga actgtctgtc cacaccccat 900  
 cccccaagc agaacctcta agccctaaa ctgtctcagac agagggtggcc cccagaacca 960  
 ggcctgcccc cctacggcat cccccagag cccaggccctc tggcacagag ccccccctcac 1020  
 ccacgcactc tttagggagaa ggtggctct tcgtgcatac ccctcagcca cgaaggccaa 1080  
 gttcccgagg ttggggccagt cccccaggtag cagggagacg ccctgtatctt ttcccttcgg 1140  
 tccctgggg cccggggccac cagggccaag ggccttgggg aacggggggg actcctcagc 1200  
 ggccccggct gtagccgtac ctcagccacc cggggccctc gtggcccttg ctgaccaacg 1260  
 gccccatgc cagccccctc tggggctct ttgtccccag tagccctatt ccaagatgtt 1320  
 ctggggagag tgaacagcta agagcctgca gccaagccgc ctgccccctc gaggcagccg 1380  
 acccccgggc ctcgcagtgc gcagcctta actccccagga attcatggc cagctgtatc 1440  
 agtgggagcc ttcaactgaa gtccagggtc cccagcgtg tgaactgaac tgccggcccc 1500  
 gtggcttccg ttctatgtc cgtcacactg aaaagggtcca ggtatgggacc ctgtgtcagc 1560  
 ctggagccccc tgacatctgt gtggctggac gtgtctgtag cccgggtgt gatgggatcc 1620  
 ttggctctgg caggcgtctc gatggctgtg gagtctgtgg ggggtgtatg tctacctgtc 1680  
 gccttgggg gggggacccctc actgaccgg gggggccctt ggggtatctag aagatcttgt 1740  
 ggattccagc gggggccctt cggctccaga ttggccagct cggccctagc tccaaactacc 1800  
 tggcacttcg tggccctggg ggccggtcca tcatcaatgg gaactgggct gtggatcccc 1860  
 ctgggtctta caggggccggc gggaccgtct ttccatataa ccgtccccc agggaggagg 1920  
 gcaaaggggg gagtctgtc gctgaaggcc ccaccaccca gcctgtggat gtotatatga 1980  
 tcttcagga gaaaaaccca ggcgtttttt atcgtatgt catctttca cctcctccaa 2040  
 tccttgagaa cccccccccc gageccccctg tccccccagct tcagccggag attctgaggg 2100  
 tggagccccc acttgcctcg gcacccggcc cagccggac cccaggcacc ctccagcgtc 2160  
 aggtgcggat cccccagatg cccgcggcc cccatccccag gacccccctg ggggtctccag 2220  
 ctgcgtactg gaaacggatg gggacactctg catgtctcgc gtcctgggg aaagggtgtc 2280  
 ggcggcccat ttccctctgc atctccctgt agtggggaga ggaactggat gaacgcagct 2340  
 gtgcggccgg tgccaggccc ccagctccc ctgaaccctg ccacggcacc ccatggcccc 2400  
 catactgggaa ggttggcggag tggacatctt gcagccgctc ctgtggcccc ggcacccagc 2460  
 accggccagct gcagtgccgg caggaattt ggggggggtgg ctccctcggtg ccccccggagc 2520  
 gctgtggaca ttccccccggg cccaaacatca cccagcttctt ccagctgcgc ctctgtggcc 2580  
 attgggaatg tggctctctt tggggccaggt gtccctgtgg gtgggggggg ggcagagaa 2640  
 gcccggcagg tggctgttt gggaaaacacg gtgtatgaatg gagcggcagg gagtgtgcgt 2700  
 caggcccccc acagcccccc acagagagg cttgtgacat gggggccctgt actactgcct 2760  
 gtttccacag cgactggagc tccaaagtct cagccggatgt tgggacggga atccagccgc 2820  
 gctctgtgtt ctgccttggg agtggggcag ccactcgggc caggccagg ggaagcagga 2880  
 gcagaactg ggcagagctg tccaaacacca agccggccccc ctgacatgcg cgcctgcagc 2940  
 ctggggccct gtgagagaac ttggcgctgg tacacaggcc cctgggtgtc gtgtccctcc 3000  
 gaatgtggct ctggcacaca gctgttagagac atcatctgt tatccaaact gggggacggag 3060  
 ttcaacgtga ttctcccgag caactgttct cacctccccca ggccccctgc cctgcagcccc 3120  
 tgtcaaggggc aggccctgcaca ggaccgtgg tttecacgc cttggagcccc atgttctcgc 3180  
 tcctgccaag ggggaacgca gacacgggg gttccagtgc tgagcaccac ccagaccctc 3240  
 agcacccgat gcccctctca actgcggccc tccaggaaacg gcccctgtaa cagccaaaccc 3300  
 tgcagccagc gcccctgtatg tcaatgcac gacagctctc cacattggcc cctgggtgt 3360  
 caggccccggc tctgtctata cccctactac acagccaccc gttggccctc ttgacggatc 3420  
 gtcctggagc ggttccctca ggatccctcc tggaaagggtt ccggggcacc ttacacggtt 3480

tctgtgcac	catcggtcac	.ccattgatcg	gcccaactctg	aaccccctgg	ctctccagcc	3540
tgtcccgagt	tcagcgaggga	tgtcctccag	gtgacagagg	gtggcaaggt	gactgacaca	3600
aagtgacttt	cagggtgtg	gtcaggccca	tgtgggtgt	tgatgggtgt	gtgcacatat	3660
gcctcagggt	tgcttttggg	actgcatgga	tatgtgtgt	ctcaaacgtg	tatcactttt	3720
aaaaaaaggagg	ttacacagac	tgagaaggac	aagacctgtt	tccttgagac	ttccttaggt	3780
ggaaaaggaaa	gcaagtctgc	agttcctg	taatctgagc	tacttaga	gtggctccc	3840
caccaactcc	agttttgtgc	cctaagcctc	atttctcatg	ttcagaccc	acatcttcta	3900
agccgcctg	tgtctctgac	cccttctcat	ttgcttaga	tcttcgcccc	tgcctcccta	3960
attagctagg	gctggggtca	gcccactgcca	atccctgcctt	actcggaa	gcaggaggaa	4020
agagactgcc	tctccagagc	aaggcccagc	tgggcagagg	gtaaaaaaaga	gaaatgtgag	4080
catcgctcc	cccaccaccc	cggccagccc	ctagccccac	tccctgcctc	ctgaaatgtt	4140
tcccacccag	aactaattt	ttttttatta	aagatggtca	tgacaaaatga	aaaaaaaaaa	4200
aaaaaaaaaa	aaaaacaaaa	aaaaaaaaata				4230

<210> 34  
<211> 3699  
<212> DNA  
<213> *Homo sapiens*

<220>  
<221> misc\_feature  
<223> Incyte ID No: 3473847CB1

tggggcagcc tcttgcctg gaactcaactc actctgggtg ctcctccccc aggtggaggt 2640  
 gccaggaagc tccctccctc actgtggggc atttcacccat tcaaacaggc cgagctgtgc 2700  
 tcgggtctg ccagctgctc ccaatgtgcc gatgtccgtg ggcagaatga cttttattga 2760  
 gctttgttc cgtgccaggc attcaatcc caggtctcca ccaaggaggc aggattctc 2820  
 ccatggatag gggagggggc ggtagggct gcagggacaa acatcggtgg ggggtgagt 2880  
 taaaagggtgc tcatggccct catctccagc taactgtgga gaagccctg ggggctccct 2940  
 gattaatgga ggcttaggtt tctggatggc atctagccag aggctggaga caggtgtgcc 3000  
 cctgtgttc acaggctgtc cctgggttcc ctgagccacc ttactctgc tctatgccag 3060  
 gctgtctag caacacccaa aggtggctg cggggagggca tcaccttaga ctgactcgcc 3120  
 agtgtcagt ggtgcatac cttgtctcc ccaaccggcact cactaccgg caggtacac 3180  
 attcgacccc ctacttcaca gaggagaaaa ctggaaacca gagggggcgt gcctgccaag 3240  
 ctcacacagc aggaactgag ccagaaacgc agattgggt ggctctgaag ccaagccctc 3300  
 tcttaacttca cccggctggg ctccctattt ttacgggtaa cagtggggct gggaaagggg 3360  
 acacagacca ggaagctgg tgagtgtatgg cagaacatgt cctgcaggca tggaactttt 3420  
 tccgttatcc cccaggcctg attactgtgc ctggcggaga tgcttctaag gcatggtcgg 3480  
 gggagggggc caacaactgt ccctccttga gcaccagccc caccaagca agcagacatt 3540  
 tatctttgg tctgtcttc tctgtgttcc ttatccagcc aacttttcta gacctgtttt 3600  
 gctttgtaa cttgaagata ttatctgg gttttagtgc atttttatta atatggtgac 3660  
 tttttaaat aaaaacaaac aaacgttgc ctaaaaaaaaa 3699

<210> 35  
 <211> 2410  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 3750004CB1

<400> 35  
 ctcagcagt ggcccccttcc ctccacgggc tgccccggag ctcagtccttca ccccttcggc 60  
 cgatgggcc atgaggcacc gaacggaccc gggccagaac ctctgtctct tcctgtgggc 120  
 cctgctgaac tgtggtttgg gggtcagtgc tcagggtccg ggcgagtggc cccctgggt 180  
 gtcctggacc cgctgctcca gtcctggccg gctggcgctc tctgtgcgc gccggcgctg 240  
 cctccggctt cttggggaaag aaccgtgtc gggagactcc catgagttacc gcctctgcca 300  
 gttgccagac tgccccccag gggctgtgcc ctcccgagac ctacagtgtg ccctgtacaa 360  
 tggcccccgt gtccctggccca cccagaagac ctaccagggtt gtgccttcc atggggcgcc 420  
 caaccagtgc gacccatcaact gctgtgtcc ggggcacccgc ttctaccaca gcttcggccg 480  
 cgtctggcc ggcacccgcg gcaaggccggg tgccctgggg gtctgcgtgg ctggcccgctg 540  
 ctttagcgcc ggctgtatgg ggttggggg ctgggttcc ctggaggacc gctgtggccg 600  
 ctgcggaggc gccaacgact cttgcctttt cttgcacccgcg gtgttgcgt acggccgtgc 660  
 cttcgctggg tactggaaacg tgaccctgtatggcc gccagacaca tccgcgttga 720  
 acacaggagc cgcaaccacc tgggtatcc aggtcaactg atggggggcg atggggcgcta 780  
 cgtgcttaat gggcaactggg tggtcagccc accaggggacc tacgaggggg cggcacgc 840  
 tgggtctac accccggagaca caggggccca ggagacattt caagcagccg gcccaccc 900  
 ccatgacctg ctccatccagc tcctctgtca ggagcccaac cctggcatacg agttttagtt 960  
 ctggctccctt cggggcgctg acagccctt ccaaggctgtgtgcggcc tggctggcc 1020  
 cctgaggccg ctcagcccccc ggggggtggaa gcttcagccc ccccgagccctgctgtc 1080  
 ccctgcacag accccaaacgc tggccccaga cccctgccc ccctgcctg acaccggccg 1140  
 cccgcggccac cgactactcc actattgcgg cagtgtactt gtgttccagg cccgagtgtc 1200  
 gggccaccac caccaggccc aggagaccgc ctatgagggtg cgcacccacgc tcgttctacaa 1260  
 gaaccgcctg ccactgcggg cacgcgagta cgtgtggccg ccaggccact gcccctgccc 1320  
 gatgtggccccc cccaccggg actacccgtatggctgtccag cgttgcgtca gccccgacgg 1380  
 cacacaggac cagctgtctgc tggcccacgc cggctacccgc cggccctggc gcccgtgg 1440  
 ggacagccgc atacgcctgtca ctggccggc ctgtctggc tgagccctg caggagcccc 1500  
 gggccacacac agcaagaaaat atacatctga ccacccctaa cgtcaacgtatggccctt 1560  
 caccctggct tccaggcagc tctgaaatac gttccacccgt tgcaagctatg tgactccctc 1620  
 ccacacacgc ttaagacacc tctgcatac gtcacccggca ctgtcacaag cccgcaggca 1680  
 ctgggtgagga ggcactaagg agactctgc ttttatttcg cctctctctt tggtgtccag 1740  
 gaagctcatac gctatttata ctcagaaaatg ttaacgcgtc ttcttttcc ttgcgcgg 1800  
 tcacacttgc ttggagacac tgcatacgtac gacccatgaca ccctgcgtcc ctgggttaccc 1860  
 agaagatcat ctgtttactt cccagacact gtgtctgtc tgctctctgc tactcacaca 1920  
 caccctcatg tgcataaggc agacacactg tcacaaacacg gcatgccccct tagaagacat 1980  
 gccttaaccac gcaactgtaaatccat gtcacccgttcc cctttccccc tggcttaccag 2040  
 gaaactcgga gacaatcttt tcagccctcg catttctggc tggatttcca cccatcaaca 2100  
 cgtgtttgtctc ctcctttttt tttttttctg aggttagaccc tgcgtctgtca cctaggctga 2160  
 agtgcgggtgg tgcaatcatg gtcactgca gcctcaaatt cctgggtca agcgatccctc 2220

ccactcagct ccacagcagt ggaactcacf tggatcaca tgccggctaa tttaaatttt 2280  
 gtagagatgg gcttgtacgt gccaaatgtc tcactatggc tcaacatctc tgctgggtcc 2340  
 aagactgaat agatgacat gatgggtac cccttatcct tatttcagct taaaaattc 2400  
 taaaaaaaaaaaa 2410

<210> 36  
 <211> 549  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 4904126CB1

<400> 36  
 gggaggagag aaaagccatg gcccacaagg tcctgaagga gaagagaaaag cagtttatcc 60  
 gttcagtggg cgaaggatac ataaatggct tactgggtga attattggag acaagggtgc 120  
 tgagccagga agagatagag atagaaaaat gtaaaaatgc tacagttatg gataaggccc 180  
 gagcttgcgt tgactctgtt attcggaaag gggctccagc atgccaatt tgcatcacat 240  
 acatttgta agaagacagt cacctggcag ggacgctggg actctcagca ggtccaaat 300  
 ctggaaatca cttactaca caagattctc aaatagatct tccttccttag gtaatgctgt 360  
 tttaaagaa agagcattct ttgaaccgtg gttcccggtg acattaatgt tggatgtga 420  
 accacagtta aaggggctat gaagaattcc catagagtga tcatacattt ttcttttgc 480  
 aatctattct gttttgttag caactgtcaa aacagttca ctatctatgt ctacattaaa 540  
 atttgaat 549

<210> 37  
 <211> 2755  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 71268415CB1

<400> 37  
 ttgcctaggag ggtggagttc atccacttat gatataaatg tctctttta ttttttgag 60  
 gcaactttttgttccttccatcacagaaaac ccattataact tcaagttgtt accctcaaac 120  
 caccacaeaggaaaatggagg abctactgtt ttaccacggc acggtgagg agacagaact 180  
 gtcacgcgtc acgctcagca cttggcagg aatttagagga ctgattacgg tgacgacaa 240  
 cctcagctacgtcatacgatc cccctccctga cagcaaggcc caacaccta ttacatgtc 300  
 tgaacatctc aagccgcccc tggggaaactg tgggttcgag cactccaagg ccaccaccc 360  
 ggactgggctt ctteagttt cacaacacac caagaagcga cctcgcaggta gaaaaaggaa 420  
 agatttaaac tccatgaagt atgtggagct ttacctcggt gctgattttagatgttca 480  
 gaagaatcga cgagaccagg acgccaccaa acacaagctc atagatgtc ccaactatgt 540  
 tgataatgtt taccgatcct tgaacatccg gattgtctc gtgggttgg aagtgtggac 600  
 ccacggaaac atgtgtgaat tttcagagaa tccatattct accctctgtt cctttctcag 660  
 ttggaggcgc aagctgtttt cccagaagta ccatgacaac gcccattaa tcacgggcat 720  
 gtcctccac ggcaccacca tcggcctggc cccctctatg gccatgtgtct ctgtgtacca 780  
 gtctggagga gtcaacatgg accactccg gaatgccatt ggcgtggctg ccacccatggc 840  
 ccacgagatg ggcaccaact ttggcatgac ccatgattct gcatgttgc gctggccag 900  
 tgccgtgtat ggtgggtcata tcattggcagc tgccactggg cacccttcc ccaaagtgtt 960  
 caatggatgc aacaggaggg agctggacag gtatctgcag tcagggtgtg gaatgtgtct 1020  
 ctccaaatcg ccacacacca ggatgttgc tgaggccgg aggtgtggg acgggtatct 1080  
 ggaagatggg gaagatgtg actgtggaga agaagagaa tgtaacaacc cctgctgcaa 1140  
 tgcctctaat tggatcctgaa ggcggggggc ggagtgtgtc cacggctct gctggccacca 1200  
 gtgttaagctg ttggctcctg ggaccctgtt cccgcgcac gcccaggatgtgtc 1260  
 ggagttctgt acggcaagt ctccccactg cccttaccaac ttctaccaga tggatgttac 1320  
 cccctgttag ggcggccagg cctactgcta caacggcatg tgcctcacct accaggagca 1380  
 gtgccagcag ctgtggggac cggagcccg acctgcccct gacctctgtc tcgagaaggt 1440  
 gaatgtggca ggagacaccc ttggaaactg tgaaaaggac atgaatgtg aacacaggaa 1500  
 gtgcaacatg agagatgcga agtgtggaa gatccagtgt cagacgtctg aggccccggc 1560  
 cctggagtcc aacgcgggtgc ccattgacac cactatcatc atgaatggga ggcagatcca 1620  
 gtgccccggc accccacgtct accgggtcc tgaggaggag ggtgacatgc tggacccagg 1680  
 gctgggtatg actggaaaccat gatgtggctt caacatatt tgctttgagg ggcagtgca 1740  
 gaacacctcc ttctttgaaa ctgaaggctg tgggaaaagtgcaatggcc atgggggtctg 1800  
 taacaacaac cagaactgccc actgcctgccc gggctggcc ccccttct gcaacacacc 1860

ggggccacggg	ggcagtatcg	acagtgggcc	tatgccccct	gagagtgtgg	gtcctgtgg	1920
agctggagtg	ttggtggcca	tcttggtgc	ggcggtcc	atgctgatgt	actactgctg	1980
cagacagaac	aacaactag	gccaactaa	gccctcagct	ctcccttcca	agctgaggca	2040
acatgttcgt	tgtccctca	gggtttctca	gaacagcggg	actggatcatg	ccaaaccac	2100
tttcaaggtc	cagacggccc	agggcaagcg	aaagggtgatc	aaactccgg	aaatccgtcg	2160
gaagccctcc	cagcctcctc	ccggggccccc	tccagattat	ctgcgtgttgc	ggtccccacc	2220
tgcaccactg	ccagctcacc	tgagcagggc	tgcttaggaac	tccccagggc	ccgggtctca	2280
aatagagagg	acggagtcgt	ccaggaggcc	tcctccaagc	cggccaattc	ccccccgacc	2340
aaattgcata	gttcccagg	acttctccag	gcctcgcccg	ccccagaagg	cactccggc	2400
aaacccagtg	ccaggccgc	ggagccctcc	cagggcagga	ggtgtcatccc	cactgcggcc	2460
ccctgggtct	ggccctcagc	agtcccgcc	tctggcagca	cttggcccaa	aggtgagtcc	2520
acgggaagcc	ctcaagggtg	aagctggta	catagggctc	caggggggca	ggtgttagagt	2580
tgagaaaaca	aagaattca	tgcttctgt	ggtctggact	gaacttccag	aacaaaagcc	2640
aagggcaaaa	cattcatgtt	tcttggtgcc	cgcttgactg	tggagtttttgc	gttcatgtg	2700
aaaggtgatt	cttagaaatcc	tgagctgtgg	tggcttcagt	cctgccccctg	cacct	2755

<210> 38  
<211> 2553  
<212> DNA  
<213> Homo sapien

<220>  
<221> misc\_feature  
<223> Incyte ID No: 7473301CB1

<400>	38	atggacaaaag	aaaacagcga	tgtttcagcc	gcacctgctg	acctgaaaat	atccaatata	60
tcagtccaa	ggttcagtgc	ccaaaagaag	ctggcagtga	gacgaccacc	gttgcaggg	120		
agacgactac	cattgccagg	aagacgacca	ccacaagac	ccattggcaa	agccaaaccc	180		
aagaagaat	ccaagaaaaa	agttccctt	tggaatgtac	aaaataaaat	catttcttc	240		
acagtat	tattcatct	agcagtata	gcctggacac	ttctgtggct	gtatatactg	300		
aagacagaaa	gcaagatgc	tttttactt	gtctggatgt	tcgttcatc	caacatcgag	360		
tttcttccg	aataaccgaca	aaaggagttc	aggaaatttc	tttcagtgtc	acggactgtg	420		
cagcaagtga	taaacctggt	ttataacaaca	tctgccttct	ccaaatttta	tgagcagtct	480		
gttgttgcag	atgtcagcag	caacaacaaa	ggcggcctcc	ttgtccactt	ttggattgtt	540		
tttgtcatgc	cacgtgccaa	aggccacatc	ttctgtgaag	actgtgttgc	cgccatctt	600		
aaggactcca	tccagacaag	catcataaac	cgAACCTCTG	tggggagctt	gcagggactg	660		
gctgtggaca	tggactctgt	ggtaactaaat	gggtattgtt	ggtcatttct	aaaaaaaaag	720		
aaaagaaaagg	aaaatgttgc	tgttccaca	gacaaaggct	gtcttcagta	tttctatgca	780		
gagcatctgt	ctcttccacta	cccgcgtggag	attttcgag	cctcaggagg	gtctgtgtt	840		
cacttcaagc	tggtgccat	agtggttctac	ctgattcgtc	tctcaatcaa	gtccatccaa	900		
atcgaagccg	acaactgtgt	cactgactcc	ctgaccattt	acgactccct	tttgcacatc	960		
cgtagcagca	tcttgtacag	aattttgtaa	cccacaagaa	cattaatgtc	atttttttct	1020		
acaaataatc	tcatgttggt	gacatttaag	tctccctata	tacggaggtt	ctcaggaatc	1080		
cgggcattatt	ttgaggtcat	tccagaacaa	aagtgtgaaa	acacagtgtt	gttcaaagac	1140		
atcaactgt	ttgaaggggaa	aatttcaagc	ccatattacc	cgagactacta	tcctccaaaa	1200		
tgcaagtgt	cctggaaatt	tcagacttct	ctatcaactc	ttggcatagc	actgaaattc	1260		
tataactatt	caataaccaa	aaagagttat	aaaggctgt	acgatgttgc	gtgggaaaatt	1320		
tatgagcaca	tgtactgtgg	ctcctacatg	gatcatcaga	caatttttcg	agtccccagc	1380		
cctctggttc	acattcagct	ccagtgcagt	tcaaggctt	caggcaagcc	acttttggca	1440		
gaatatggca	gttacaacat	cagtcaaccc	tgcctgtgg	gatcttttg	atgtccctcc	1500		
ggtttatgt	tccctcaggc	ccagcgtgtt	gatggagtaa	atgactgttt	tgtatgaaagt	1560		
gatgaactgt	tttgcgttag	ccctcaacct	gcctgcata	ccagctccct	caggcagcat	1620		
ggcccttc	tctgtatgg	cttcagggac	tgtgagaatg	gccccggatga	gcaaaactgc	1680		
actcaaagta	ttccatgcaa	caacagaact	ttaaagtgt	gcaatgtat	ttgcctttag	1740		
aaacaaaatg	caaaaatgtga	tgggacagt	gatttgcag	atggaagtgt	tgaagaaggc	1800		
tgcacctgca	gcaggagttc	ctccgcctt	caccccatca	tcggaggcac	agacaccctg	1860		
gaggggggtt	ggccgtggca	ggtcagcctc	cactttgtt	gatctgccta	ctgtggtgcc	1920		
tcagtcatct	ccagggagtg	gtttcttct	gcagccact	gtttcatgg	aaacaggctg	1980		
tcagatccca	caccatggac	tgcacaccc	gggatgtatg	ttcaggggaa	tgccaagttt	2040		
gtctcccccgg	tgagaagaat	tgtgttccac	gagtaactata	acagtcaagac	ttttgattat	2100		
gatattgtt	tgtacatagct	cagtattgcc	tggcctgaga	ccctgaaaca	gtcatttcag	2160		
ccatatatgca	ttcctccac	tgttcagaga	gttgcagt	ggggagaatg	ctgggttaact	2220		
ggctgggggc	gaagacacga	agcagataat	aaaggctccc	tcgttctgc	gcaagcggag	2280		
gttagagctca	ttgtatcaaac	gctctgtt	tccacactag	ggatcatc	ttctcgatg	2340		
ctctgtgcag	gcataatgtc	aggcaagaga	gatgcctga	aaggagattc	gggtggaccc	2400		
ttatcttgc	gaagaaaaaaag	tgatggaaaa	tggatttga	ctggcattgt	tagctgggaa	2460		

catggatgt gacgacccaa ctttctgg ttttacacaa gggtgtcaaa ctttgc 2520  
 tggattcata aatatgtccc ttctctttt taa 2553

<210> 39  
 <211> 1041  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 7473308CB1

<400> 39  
 atgttcagcg gcaacacagg aaaaacccat attatcaatg ctcaaaaaacc tggcccaccc 60  
 aggcttagcc agttattcgt gagcagagag gtgtgtcatc tacatggcag tcattggcctg 120  
 gatgggtctg gaactgtggc aagaatccctt ccagggaaaca gcccgtotcc ctctctgctc 180  
 tcagaaggca agtttcctta tcacctgtct gcttcagaa ggcaagtttc cttatcacct 240  
 gtgaatcaca aacccacacaga gtggccaaac atactgtatc aagaccatag gaaggggaaa 300  
 gctgcagttg gtgtctcctt tgatgtatc gacaagattt ttgggggcta caactgtgag 360  
 gagaattctg tccccatcca ggtgtccctg aattctggct accacttctg ttttggctcc 420  
 ctcaacaggg aatactgtatc ccaggtgaga ctgggagagc acaacatcga agtccctagag 480  
 gggaaatgtaaac agttcatctt tgccgtcaag atcatccgcc accccaaataa caacagctgg 540  
 actctggaca atgacatcct gctgtatcga ctctccacac ctgcacatcat caatggccat 600  
 gtgtccacca tctcttcgtt caccacccct ccacgtctg gcaactgtatc cttatctct 660  
 ggctggggca acactctgag ttctggccg gactacccag acgagctgca gtgcctggat 720  
 gtcctgtgc tgagccaggc tgagtatgaa gcctcttacc ctggaaagat taccacaaac 780  
 gtgttttgtg tgggtttcct tgagggaggc aaggatttctt gccagattat tcctatcaaa 840  
 gtgcagcagc tggttaccc aagccaaagag acagacataa ggatccctat ggcccttgcag 900  
 acagctgtttt ccacccatcca cctggggccc ttagactt tacacaggaa agtgagtcc 960  
 :cccaactgaga; agcgttgcca gcagaaacag ggcatgaaaa tcacagataa ccatgggatt 1020  
 :cttccaagt ggtcagtata a 1041

<210> 40

<211> 117.07

<212> DNA

<213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 7478021CB1

<400> 40  
 matgtctccggc cctccatctt ccgtccgaca ctgtgtctt gctggctggc tgctccctgg 60  
 cccaaacccagg cccggaggatct cttccacacgc cgggaccgct cggacccatgg gccgtccccc 120  
 ctgcgcceagg ccaagcccat tgccgacccct caccgtctc agccgttctt gtcaggatata 180  
 ggctgtctcgg ggggtgtgggg ggcctggggg cccagttcccg agggggccccc ggagacccccc 240  
 aaggggccggc cctctggccaa ggcgggtcgcc aggtttccagc gggcgaacgc gtcggccggcc 300  
 agcggggaggc tggacgcggc caccctatcg gcatgaaacc ggccggcgctg cgggggtcccg 360  
 gacatgcgcc caccggccccc ctccggccccc ctccggccccc cggggccccc cccagagcc 420  
 cgctccaggc gtcctcccgcc ggcggccgtg tccttgcctt ggcgggggttgc gacggccccc 480  
 ggctaccccg acggggggagc tgcccaggcc ttctccaaga ggacgtctgag ctggccgtct 540  
 ctggggcagg ccctgagccg ccaactgtcc gtggccgacc agcggccgt agaggccgt 600  
 gccttcaggc tggtggagccg ggtgaccccg ctggacttcc gcgaggaccc ggcggccccc 660  
 gggggccggc tgcacatcaa gctgggtttt gggagacggc acctgggtctg tccggggccc 720  
 ttctcgatggg gccggggaggc gtttgcacac gctggccccc taggtgacat tcactttgac 780  
 gacgacggcc acttccacacc tcccacccgt gacacgggca ttcggccctt caaggtggcc 840  
 gtccatgaaa ttggccatgt cctgggtttt ctcacaccc acaggacggg atccataatg 900  
 caaccaaatt acatcccca ggagccgtcc ttggatgttgg actggtcaga cagggaaagca 960  
 attcaaaaggc tttatgtggttc ctgtgggggatcatttgcata ctgcgtttga ctggatttgc 1020  
 aaagagagaa accaatatgg agaggtgtatc gtggatgttgc acatattt cttccgtaaac 1080  
 agctgtact ggcttttatga aaatcgaaac aataggacac gctatggggc ccctatccaa 1140  
 atccctactg gtcggccctgg aatcccaaca cacaacatag atgcctttgt tcacatctgg 1200  
 acatggaaaa gagatgtaaac ttatttttt caagggaaatc aatactggg atatgacagt 1260  
 gacaaggatc agggccctac agaagatgaa caagggaaaaa gctatccaa attgatttca 1320  
 gaaggatttc ctggcatccc aagtccctta gacacggcgt ttatgaccg aagacagaag 1380  
 ttaatttact tttcaagggat gtccttgcata ttggatgttgc atgtcaacag aaatcgatg 1440  
 ctttttttttccat atccaaagag gattactgaa gtttttccag cagtaatacc acaaaaatcat 1500

ccttcagaa atatagattc cgcttattac tcctatgcat acaactccat tttcttttc 1560  
 aaaggcaatg catactggaa ggttagttat gacaaggaca aacaacagaa ttctgggtt 1620  
 cctgtaatg gtttattcc aaaaaagttt atttcagaga agtggttga tggttgac 1680  
 gtccatatct ccacactgaa catgtaa 1707

<210> 41  
 <211> 1262  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 4333459CB1

<400> 41  
 aaaagatctt tgcgaaaacac tacattcaga aacatcgat ggacatgctt gattcaccac 60  
 gtctggta atgaataaac ttgtttaaa ttggcttatt gctggctctt caaggcttcc 120  
 tattttgtt tgcttttagtc tctctaaaat ttcaaggaaa aactatgagt ctcaaaatgc 180  
 ttataagcag gaacaagctg attttactac taggaatagt ctttttgaa cgaggtaaat 240  
 ctgcaactct ttgcgtcccc aaagctccc ttgtgtggca gagtctggtt aaggtaacgc 300  
 cttggaatta tttaacatt ttctcgtcga ttcttggagg aagccaagtg gagaagggtt 360  
 cctatccctg gcaggttatot ctgaaaacaaa ggcagaagca tatttggta ggaagcatcg 420  
 tctcaccaca gtgggtgatc acggcggctc actgcatttc aaacagaaaac atttgttcta 480  
 ctttgaatgt tactgtctgaa gatgtact taaggccagac agacccagga gagcaaactc 540  
 tcactattga aactgttcatc atacatccac atttctccac caagaaacca atgactatg 600  
 atattccct tttgaagatg gctggagcct tccaatttgg ccacttggt gggcccat 660  
 gtcttccaga gctgcgggg caatttgagg ctggtttat ttgtacaact gcaggctggg 720  
 gcccgttaac tgaagggtggc gtccttcac aagtcttgc ggaagtgtaat ctgcctattt 780  
 tgacctggga agagtgtgtg gcagctctgt taacactaaa gaggccatc agtgggaaga 840  
 ccttctttc cacaggtttt cctgtatggag ggagagacgc atgtcaggga gattcaggag 900  
 gttcactcat gtgcgggaaat aagaaagggg cctggactct ggctgggtgt acttcctggg 960  
 gtttggctg ttgtcgaggc ttggaaaaca atgtgaggaa aagtgtatcaa ggatcccctg 1020  
 ggatcttcac agacattatg aaagtgtttt cctggatcca cgaacacatc caaactggta 1080  
 actaagccat cacacaaggt taagaagctg ccattctgtc agggccagag acagcatcag 1140  
 cagagtctg gcaaatcaga gcacctgaac caacaggctc tacctctgtt ctcaagtgt 1200  
 cacacaagga ttgtgagggtt taccagtct aaataaaaaca agagttaaat atgtaaaaaa 1260  
 aa 1262

<210> 42  
 <211> 3067  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 6817347CB1

<400> 42  
 gcactgtgaa cgttgggtgc atccaaatct gaattttgtc tgggaccagg gtcaggacc 60  
 agaatacacc agagctgagg gccagcccta cctgagaacc atcaacaaaac ttaccccaca 120  
 tcccattata cctcctcaact ccctgcagcc tgcagttt ccaatctcc cacactca 180  
 gtcacctggg gctctgggtgc accagatgac actacttgc ccctggtaca cagggccat 240  
 gatccccatg gatgttaatg agcccaactc cgtgaccacg gctccttaccc tcagctctag 300  
 cctgcagcat atctcctcat tcctggccac ttgtggaaaat cttccctcc attttggtca 360  
 tccacgttag tttgtcaatggc ccaggattga tgaaaaaat agaaggat ttggagacag 420  
 tgagccaggt gccaaactct tcaataatgt tgagttctgt ttgtgcctgc aaaaacgggg 480  
 gccagtgaac attacatca gttgtgttag tcggcggacc ttacaaatat cagttttgt 540  
 gttatcagag aaatacgagg gtattgttaa atttgaatcg gatgaattac ctttgggtgt 600  
 aattgggtct aatattgggt atgcacattt tcaagaattc agggctggaa tctcctggaa 660  
 gcctgtggta gatcctgtatg accccattcc tcagttccct gattgtcga gcagcagcag 720  
 cagcaggatt cttcgtgtatg ttgtgttagt tcagttccct ctgggtcag gccacaaaagg 780  
 gcaggcattt attgtaaaaggc ttgtgggtg ctcaaggaa ttgaagcaag agctgactca 840  
 ggaaggggccg ggcgggggac accccaggtc tgcgtggccc cccgcggcc acggccactg 900  
 gcccggccgag ccctgcgacg aggggggaga gcccggccca gtggaggccg aggaggtaga 960  
 ggaggccgag acggccggaga aggccggagag gaaggtggag gcccggccga aggtggaggg 1020  
 gaagggccgag gcccggggggc agggcggagc gcccggggaa gtggacgcca ccgagaagg 1080  
 ggagacggcg gggaaagggtgg acggccgtgg gaaggtggag acggccggagg gtccggccg 1140

ccgggcttag ctcagaatgg agcccaacc cgagccgtc cgggaggcg 1200  
 gaaggcaggag ctggaggatg agaaccacgc gcccggcgc ggtggcgca acacgcacga 1260  
 gttctccccc cccacccttc cctccatcc acccgccccc cccgatccct ctccgcgtcg 1320  
 cagtctgtcg cccgcggcc gacccggcc cccggccccc accccggctcc gtaccccgcc 1380  
 gcagctttagg ccccgcccccc cccccggcc cccggccccc cccggccctg gggcgatg 1440  
 cctggatgtg gattttgcgg tggggccacc aggtgttct cactgtaaaca gcttaaggt 1500  
 gggagagaac tgaggcagg aactgcgggt tatctaccag tgcttcgtgt ggtgtggaaac 1560  
 cccagagacc agaaaaagca aggcaaagtctgcatctgc catgtgtg gcacccatct 1620  
 gaacagactc cacttttgc ttccctgtgt ctttttggc tgcttcacgg agaaaacacat 1680  
 tcacgacac goagagacaa aacaacacaa ctttagcaga gacctgtatt acggaggat 1740  
 atactgtttt atgtgttaagg actatgtata tgacaaaagac attgagcaaa ttgccaaaga 1800  
 agagcaagga gaagcttga aattacaacgc ctccaccta acagagggtt ctcaccagca 1860  
 gtgttcagtgc cccggcttg gtgagaaatt cccaaacctgg gaaacaacca aaccagaatt 1920  
 agaactgctg gggcacaacc cggaggaaag aagaatacc tccagctta cgatcggtt 1980  
 aaggagactc atcaatcttgc gcaacacgtg ctttatgaac tgcattgtcc agggccctcac 2040  
 ccacaccccg atactgagag attctttct ctctgacagg cacccatgtg agatgccgag 2100  
 tcccggatgtg tgcgttgtt gtgagatgtc gtcgtgtt cggggatgtt attctggaaa 2160  
 cccgtctccatgtgcctt ataaatgttgc gacacgtgtt tggtacatgt cccgcattt 2220  
 agcagggtac aggcaacacagg atgcccacga gttccttatt gcacgtttag atgtctgc 2280  
 caggcactgc aaagggtatg atgtcgggaa ggcggccaaac aatcccaacc actgttaactg 2340  
 catcatagac caaatctca caggtggct gcagtcgtat gtcacctgtc aaggctgcca 2400  
 tggcgtctcc accacgtatg acccatgtgc ggacattgtt ttggacttgc ctggctttt 2460  
 caccccttc tggccatgtc gcccaggggag ggagggatgtt gtgaacgggg aaagccacat 2520  
 accaggaatc accaccctca cggactgtt gggggatgtt acggggccag agcactttagg 2580  
 aagcaatgtcc aaaatcaat gtggtagttt ccaaaagctac caggaatcta ccaaacagct 2640  
 cacaatgtat aatttacatgc tcgttgcctt tttcatttc aaacgggtt aacattcagc 2700  
 gaaacagagg cgcaagatca ctacatacat ttcccttctt ctggagctgg atatgacgcc 2760  
 gtttatggcc tcaagataaatg agagcagaat gaatggacaa ttgcagctgc caaccaatag 2820  
 tggaaacaac gaaaataatgtt attcctgtt tgctgtgtt aatcaccacaa gAACCTTGA 2880  
 gagtgccacataaccatgtt tcatccggca ccacaaggac cagtggttca agtgtatgt 2940  
 tgccgtcatc actaaggccatgtt tgatgtgtt aatcaccacaa gAACCTTGA 3000  
 ctatcacaaa caggtgtcttag aacatgagtc agaaaaatgtt aaagaaatgtt acacacaacg 3060  
 ctatcataatgtt 3067

THIS PAGE BLANK (USPTO)

**This Page is Inserted by IFW Indexing and Scanning  
Operations and is not part of the Official Record**

## **BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- BLACK BORDERS**
- IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**
- FADED TEXT OR DRAWING**
- BLURRED OR ILLEGIBLE TEXT OR DRAWING**
- SKEWED/SLANTED IMAGES**
- COLOR OR BLACK AND WHITE PHOTOGRAPHS**
- GRAY SCALE DOCUMENTS**
- LINES OR MARKS ON ORIGINAL DOCUMENT**
- REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**
- OTHER:** \_\_\_\_\_

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.**

*THIS PAGE BLANK (USPTO)*